# Vascular plants in modern agricultural landscapes of SE Norway: Spatial distribution and temporal dynamics

**Anette Edvardsen**


**Dissertation presented for the degree of Philosophiae Doctor**

**Natural History Museum**

**Faculty of Mathematics and Natural Sciences**

**University of Oslo**

**2010**

*Series of dissertations submitted to the*
*Faculty of Mathematics and Natural Sciences, University of Oslo*
*No. 1004*

ISSN 1501-7710



Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

*Do what you like,*
*like what you do.*

# Takk

Dette arbeidet har blitt utført ved Seksjon for forskning og samlinger, Naturhistorisk museum (NHM), Universitetet i Oslo i tidsrommet 2006-2010.

Først og fremst vil jeg takke veilederne mine, Rune Halvorsen, Ann Norderhaug og Knut Rydgren for godt samarbeid og faglig påfyll. Ann har vist frem de vakreste kulturmarksenger, og hun har tatt meg med på en reise tilbake i tid for å forstå hvordan kulturlandskapet og dets endringer i bruksmåte påvirker artene som finnes der. Ingenting er som å komme til Vestlandets perle, Sogndal! Knut er gourmetiker og disker opp med de herligeste retter, mens han villig vekk forteller om steintippenes hemmeligheter og mosenes populasjonsbiologi. Vegar har ingen formell veilederrolle, men stiller alltid opp og har villig bidratt når GIS-utfordringene har blitt for store.

Takk til alle i IMN-gruppa ved NHM for mange interessante faglige diskusjoner og også innspill på resultater og planlegging av feltarbeid! Takk til 3Q-gruppa ved Skog og Landskap for hjelp og opplysninger når det trengtes. Både Marianne og Anders har vært de beste rom-kamerater man kan tenkes å dele kontor med; rolige, ryddige og alltid med et godt svar på lur. Inger og Liv Norunn i Sogndal har også vært gode arbeidskolleger til tross for at vi jobber på geografisk adskilte steder; dere har vært et veldig hyggelig bekjentskap, i tillegg til å være nyttige diskusjonspartnere angående problemstillinger knyttet til biologisk mangfold i kulturlandskapet. Ellers vil jeg også takke andre tidligere og nåværende studenter og kolleger ved NHM og Biologisk institutt for å ha bidratt til å gjøre studietiden både hyggelig og interessant. Spesiell takk rettes til Siri R. og Anne M. for at dere gjør det så hyggelig rundt dere på museet, til Wenche H. på biblioteket for at hun finner litteraturen man leter etter, til Reidar E. for innsatsen på vriene artsbestemmelser, og til Jan Ingar I. B. for å ha vært en tidlig inspirasjonskilde.

Jeg vil dessuten takke alle grunneierne som har gitt meg lov til å utforske gårdsdammene deres, annen innmark og utmark, og som villig har vist frem gamle bilder og fortalt om brukshistorikken til de forskjellige områdene. Jeg har sett mange flotte steder og feltarbeidet har gitt mange minnerike opplevelser; alt fra å kjempe seg gjennom metervis

av brennesle og tistler rundt gårdsdammene (for ikke å glemme iglene) til møte med ulv i de dype Østfoldskogene på leting etter griseblad.

Til sist vil jeg takke familie og venner for å ha vist interesse for arbeidet mitt. Mamma, broren min og tante Gro har vært med på jakt etter griseblad. Mamma har flittig lappet og vasket møkkete felt-tøy, hun har smurt de lekreste matpakker hver eneste dag mens jeg har vært på feltarbeid og bodd hjemme, i tillegg til at hun har passet på de 237 haugene med jordprøver som lå strødd utover gulvet hjemme for å tørke. Pappa har vært oppfinnsom og laget feltutstyr som trengtes, og han har passet på at den lille, røde racerbilen har vært i stand til å ta seg frem til alle avsidesliggende steder der feltarbeidet skulle gjøres. Mamma og pappa har dessuten bidratt litt økonomisk underveis i studieløpet og bør således takkes for en del av de 545 studiepoengene som har blitt avlagt i løpet av alle studieårene.

Den største takken går så klart til Rune. Som veileder er han faglig oppdatert, ryddig og rask med å gi tilbakemelding, han har alltid mange ideer og er svært kunnskapsrik. Denne mannen gjorde inntrykk allerede første gang han dukket opp i gummistøvlene sine på tur gjennom Oslos gater den 10. mai 2002, klar for tur i Østmarka. De fleste som leser dette, vet vel hvordan denne historien fortsatte. Etter at vi leste de etiske retningslinjene for veiledere høsten 2007, så var det aldri et alternativ å bytte veileder; et valg som jeg for øvrig aldri har angret på. Tusen takk for alt du har lært bort, takk for alle de fantastiske naturopplevelsene vi har hatt sammen, og takk for at du har vært så flink til å balansere rollene som veileder og verdens beste kjæreste!

Oslo, juni 2010

*Anette Edvardsen*

# Table of contents

# Abstract

This thesis addresses the spatial distribution and dynamics of vascular plant species in modern agricultural landscapes in SE Norway. This is done by analysing several data sets for species composition and environmental factors; the most important are 2201 patch elements located in 16 1-km$^2$ squares in SE Norway; data from farm ponds; and two data sets from present and former semi-natural grasslands, respectively. Eventually, results are compared to a companion study in a more traditional agricultural landscape of W Norway. Spatial and temporal patterns in species richness and composition are analysed by habitat specificity measures, ordination analyses and generalised linear modelling. A semi-natural affiliation index is used to assess the affiliation of each patch's flora to long-term extensively managed semi-natural grasslands versus intensively used agricultural land and waste ground.

This thesis also includes a theoretical evaluation of habitat specificity indices, which forms the basis for the empirical studies. The theoretical study provides arguments for use of gamma diversity contribution and core habitat specificity measures in parallel and the empirical studies show that the context used for quantifying habitat specificity strongly influences the results. High values for both measures were observed for woodland, pastures and road verges, whereas midfield islets, boundaries and ploughed land types were in general ranked low, in both modern and traditionally used landscapes. These findings were also supported by the semi-natural affiliation index. Semi-natural grassland species were scatteredly distributed throughout whole landscapes but peaked in woodlands in addition to semi-natural grasslands. Both information on past land-use derived from an old cadastral map in addition to results from the study of modern agricultural landscape patches characterised by the semi-natural species *Scorzonera humilis*, indicate that many semi-natural affiliated species persist long after management practices has come to an end, and that an extinction debt may be present in the flora of abandoned grassland patches.

Floristic gradients of agricultural landscape patch types identified by ordination methods were related to regrowth succession (reflecting long-term management and land-use intensity) and nutrient status. On the other hand, farm pond ecosystems showed few patterns in species compositional change over time, even though a slight increase in the nutrient content of pond waters was observed. The results indicate that plant species

composition in ponds will largely be determined by historical, idiosyncratic events in addition to current environmental conditions.

To acquire a better basis for assessment of the status for rare, red-listed species, mapping and monitoring is necessary. Predictive habitat distribution models, which relate the geographical distribution of a species to its environment, are shown to be a promising tool in ecology. By using spatial predictions, sampling efficiency of rare species may be increased and finding new occurrences of rare species may be simplified. The applied use of habitat suitability models has been explored by making fine-scaled spatial predictions of a red-listed species, *Scorzonera humilis*, in Norway. Use of independent evaluation data adds strongly to credibility of habitat suitability models of rare species.

## List of papers

This thesis is based on six papers which will be referred to by their corresponding Roman numerals (I-VI). The published papers I, II and VI are reprinted with kind permission of the publishers.

**I** Halvorsen, R. and Edvardsen, A. 2009. The concept of habitat specificity revisited. Landscape Ecology 24: 851-861.

**II** Edvardsen, A., Halvorsen, R., Norderhaug, A., Pedersen, O. and Rydgren, K. 2010. Habitat specificity of patches in modern agricultural landscapes. Landscape Ecology 25: 1071-1083.

**III** Edvardsen, A., Halvorsen, R. and Norderhaug, A. The present distribution of a rare, semi-natural grassland species in a modern agricultural landscape: exploitation of new habitats or the ghost of meadows past? Manuscript.

**IV** Edvardsen, A. and Halvorsen, R. Temporal and spatial dynamics of vascular plants in SE Norwegian farm ponds. Manuscript.

**V** Edvardsen, A., Bakkestuen, V. and Halvorsen, R. A fine-scaled habitat suitability model for the red-listed vascular plant, *Scorzonera humilis*. Submitted manuscript.

**VI** Hamre, L.N., Halvorsen, R., Edvardsen, A. and Rydgren, K. 2010. Plant species richness, composition and habitat specificity in a Norwegian agricultural landscape. Agriculture, Ecosystems and Environment 138: 189-196.

# Introduction

Agricultural landscapes are landscapes of change and have been influenced by humans over generations (Norderhaug et al., 1999). Conjunctures, technology, agricultural and environmental policies, in addition to environmental factors such as climate and soil conditions, will strongly affect the way we choose to utilise the landscape. During the past decades a polarisation has taken place within the agricultural landscape; already intensively managed land is subjected to increasingly intensified use while more extensively managed land is abandoned (Robinson and Sutherland, 2002; Poschlod et al., 2005). After World War II the Norwegian authorities have encouraged farmers to change from farming involving domestic animals to production of grain in those parts of the country where the natural conditions make this possible. This is especially relevant for the south-eastern part of Norway and has been supported by economical measures. Consequently, this polarisation process may lead to a less complex landscape structure. It is assumed that the simplification may reduce biodiversity at all levels (ecosystem, community, species, population and genes). Such biological changes may have wide-ranging consequences, the nature and extent of which remain largely unknown (Loreau et al., 2001).

Less than 3% (10 903 $km^2$) of the total land area in Norway is used for cultivation purposes (Strand and Bekkhus, 2008). This is a low figure compared to for instance neighbouring countries [Sweden: 6%; Finland: 8%; Denmark: 53% (Anonymous, 2009)]. However, the agricultural landscape may also be defined as agricultural land including all areas within a 100 meter buffer zone around this land type (Strand and Eriksen, 2008). According to this definition the Norwegian agricultural landscape covers (29 073 $km^2$) or 9% of the total Norwegian land area. Approximately 40% of the total Norwegian cultivated agricultural area occurs within the south-eastern part of the country, comprising the counties Østfold, Vestfold, Akershus, Hedmark and Oppland. The landscape in this region is mainly characterised by large, contiguous areas of ploughed land. In the more mountainous parts of the country and along the fjords in the western part of Norway, agriculture is typically concentrated along the more productive river plains of the valley floors. Additionally, steep hill sides facilitate animal husbandry and the landscape is in general more fine-scaled due to terrain variation (Hamre et al., 2007).

Vascular plants represent an important part of landscape biodiversity. Many of the typical species occurring in the agricultural landscape are culturally dependent, i.e. active management is crucial for their continued existence (Brys et al., 2004). To successfully manage semi-natural grasslands, knowledge of ecological processes which determine floristic variation is of importance (Økland et al., 2006; Klimek et al., 2007). Traditionally, infield patches were grazed during springtime and then mown and dried ultimo July when the livestock was situated in the outfields (Norderhaug et al, 1999). These practices were followed by late summer grazing. This created light-open semi-natural grasslands inhabited by plant species that are specifically adapted to these management practices, i.e. semi-natural grasslands species. Long continuity of management such as grazing or mowing is one of the most important factors for maintenance of high species richness in semi-natural grasslands (Cousins et al., 2007).

Small patches of remnant vegetation (e.g. midfield islets, ponds, woodlots, former hay-meadows or grazing-land) may hold a key position in intensively managed landscapes by serving as the only hospitable habitat islands for a wide range of species (Cousins and Lindborg, 2008). Linear landscape elements such as ditches, hedgerows, field margins and road verges may also function as habitats in their own right as well as by linking habitat patches together (Berge and Hestmark, 1997). Semi-natural grasslands like hay meadows were widespread across Europe less than one century ago but have declined since then (Van Dijk, 1991). This also holds true for farm ponds (Edvardsen and Økland, 2006a). Formerly, farm ponds could be the only source of water at a farm and they commonly had multiple functions there. At present modern equipment and technology like water conduit systems and easy access to fodder intended for domestic animals have made habitats like farm ponds and hay-meadows superfluous. Assisted by a strong reduction in farms in Norway (2 out of 3 holdings have closed down during the time period 1949-1999; Anonymous, 2002), many of the species typically associated with small, eutrophic water bodies and other marginal habitat patch types in the agricultural landscape are now becoming rare and red listed (Kålås et al., 2006). Consequently, knowing the variation in species composition and species richness in modern agricultural landscapes is important for appropriate biodiversity management.

**Diversity indices as tools for conservation measures and recognition of unique sites**

Biodiversity has long remained a difficult concept (Noss, 1990), notably because of the many different aspects of biological diversity that can be measured (e.g., see Whittaker,

1972; Peet, 1974; Pielou, 1975; Magurran, 1988). Ecologists commonly group species diversity measures into three categories: within-community diversity (α), between-community diversity (β), and total diversity within a region or study area (γ) (Whittaker, 1960; 1972). However, the concepts of within- and between-community diversity are complex. Within-community diversity may include richness (the number of species present within an area) and/or evenness (equitability of species in the community with respect to biomass or other aspects of abundance). Additionally, indices may also combine both richness and evenness [e.g. Shannon's (Shannon and Weaver, 1949) and Simpson's (1949) indices].

A main goal of conservation biology is to find appropriate management strategies for preservation of nature's variation. This holds for regional-landscape, community-ecosystem, population, species, and genetic levels of biological organisation. Characterising landscape biodiversity in a way that reflects conservation value has long remained a challenge (Stohlgren et al., 1997; Økland et al., 2006). Areas are often targeted for conservation purposes because of which species, rather than the number of species, it holds. As a possible solution to this challenge Wagner and Edwards (2001) proposed the concept of 'habitat specificity'. This quantitative expression reflects each patch's contribution to species richness at the landscape scale. Habitat specificity indices represent richness ($\alpha$) and/or distinctiveness ($\beta$) components of diversity. The latter may be defined by $\alpha$ and $\gamma$ (landscape) diversity in two alternative ways: multiplicatively ($\beta = \frac{\gamma}{\alpha}$) (Whittaker, 1960) and additively ($\gamma = \alpha + \beta$)(MacArthur et al., 1966; Lande, 1996).

Habitat specificity analysis thus provides a tool for partitioning landscape species diversity on landscape elements, notably by separating patches with many rare specialist species from patches with the same number of species, all of which are common generalists. Such information will thereby provide knowledge of relevance to conservation goals at regional and national levels. Species diversity indices are therefore potentially important as quantitative tools by estimating each patch's value for different conservation purposes in the landscape.

**Biodiversity assessment by use of habitat distribution modelling**

Mapping the distribution and monitoring the populations of rare species is required for management of biological diversity (Kålås et al., 2006). Using predictive habitat distribution models has become a promising tool in ecology (Guisan and Zimmermann,

2000; Guisan et al., 2006). This is done by relating the geographical distribution of a species to its environment. Such species – environment relationships are generally based on hypotheses about how environmental factors control the distribution of species. Information about population trends, e.g. by habitat suitability modeling methods, may therefore provide valuable supplementary information of status for species associated with agricultural land. This relates to the rapid shifts in agricultural land-use in addition to the large quantities of red-listed species associated with such habitats (Kålås et al., 2006).

**Persistence of semi-natural grassland species gives rise to extinction debt**

Patterns of variation in species richness and composition (commonly expressed by coenoclines; Whittaker, 1967) relates both to environmental or historical complex factors (Økland, 1990). However, land-use history may in general be difficult to measure and decode into present explanatory variables due to the large variation in land-use practices over time (Vandvik and Birks, 2002). Furthermore, this is complicated by variation in the spatial scales over which land-use practices impact the landscape. Old aerial photographs and digitised old cadastral maps, however, provide useful information on former land-use and past land-use geometry (Hamre et al., 2007).

Comparisons of present and historical landscape patterns clearly demonstrate the ongoing loss of traditionally managed and species-rich semi-natural grasslands. This process in turn leads to a decline in vascular plant species associated with such habitats (Luoto et al., 2003). The affiliation of the flora of any patch in a landscape to extensively managed semi-natural grasslands can be approached by calculating a semi-natural grassland affiliation index (SNI; Edvardsen et al., in press). This index reflects species composition in each patch or patch type, calculated as the proportion of semi-natural species compared to species associated with intensively used agricultural land and waste ground (habitats mostly dominated by ruderal species). The abundance and persistence of species affiliated with semi-natural grasslands are determined by interaction between several factors, of which habitat quality is important. Habitat quality is reduced by use of pesticides and fertiliser (Willi et al., 2005), which may subsequently facilitate nitrophilous species. Furthermore, abandonment of formerly extensively used areas is mostly accompanied by regrowth (Strijker, 2005), most often to (re-)forestation of grasslands which eventually turn into secondary woods (Hamre et al., 2007). Such abandoned habitat patches are often marginal; situated in rugged terrain and commonly not easy accessible by agricultural machinery, and thus their economical profit is low.

Nevertheless, historical land-use factors may influence species composition for several decades after direct management has come to an end (Tilman et al., 1994). The time-lagged response offers an opportunity for many species to disperse into other patch types in the modern agricultural landscape, if such patches exist. Many species typical of semi-natural grasslands have long-lived individuals that may persist for a long time after abandonment. This discrepancy between current species composition and environmental factors has been referred to as an extinction debt (Tilman et al. 1994). The extent to which the flora of agricultural landscapes of different kinds exhibits extinction debts is, however, not yet settled (Cousins 2009).

**Spatial and temporal dynamics**

The complex interacting factors portraying the cause and effect of changes in the agricultural landscape operate on both spatial and temporal scales (Farina, 1998). Spatial patterns are somewhat difficult to verify and this may also be the case with patterns reflecting different time periods. Most ecological studies of plants in agricultural landscapes have focused on selected species (e.g. Brys et al., 2004) or patch types [e.g. meadows (Losvik, 2007); field margins (Smart et al., 2002)], most often within a smaller geographical area (see Bratli et al., 2006). Assessment of biodiversity at the landscape scale has long remained a challenge (Waldhardt 2003) and few studies have addressed species distributional patterns in entire landscapes (but see Økland et al., 2006; Liira et al., 2008). In this thesis I intend to explore the patterns of variation in species composition and richness that occur in agricultural landscapes. Specifically, I examine which landscape elements contribute to landscape species diversity and the relationship between species composition/richness and environmental and land-use factors. Understanding the ecology of agricultural landscape patches and identifying the factors which contribute to spatial and temporal patterns of species composition and diversity, is of importance for the applied management and use of agricultural landscapes in the future.

## *Aims of the study*

The general purpose of this study was to examine spatial and temporal dynamics in vascular plant species composition and diversity in modern agricultural landscapes and to distinguish which factors contribute to these patterns.

More specifically I wanted:

Paper I (i) to examine the concept of habitat specificity as a family of diversity indices for ranking of patches in a landscape on the basis of their species content; (ii) to compare variants of habitat specificity indices by use of theoretical reasoning and real example data from modern agricultural landscapes; and (iii) to discuss the practical usefulness of different habitat specificity indices.

Paper II (i) to examine if rare species tend to co-occur in the same types of habitat patches in the modern agricultural landscape; (ii) to identify the land types that hold the most habitat specific flora; and (iii) to search for relationships between habitat specificity and the characteristics of single landscape elements such as area, shape and basic ecological properties.

Paper III (i) to explore how intensification or abandonment processes in modern agricultural landscapes influence species composition in semi-natural grasslands by using *Scorzonera humilis* as a characteristic species of extensively managed grasslands; will the species exploit new habitats or is it restricted to its original habitat?; (ii) to discuss the relevance of these results for the broader evaluation of the future prospects for semi-natural grassland species in modern agricultural landscapes.

Paper IV (i) to study vegetational changes in 63 ponds in the SE Norwegian agricultural landscape over a six-year period; and (ii) to discuss factors that may be decisive for the observed patterns.

Paper V (i) to explore the applied use of habitat suitability modelling as a tool for making spatial predictions of occurrences of a red-listed vascular plant species associated with semi-natural grasslands in Norway, *Scorzonera humilis*, on a fine scale using accurate

georeferenced data for all records; and (ii) to discuss the use of independent evaluation data for evaluation of such models.

Paper VI (i) to examine variation in, and factors that affect, overall patterns of vascular plant species richness, species composition and habitat specificity among patches and patch types in an agricultural landscape created by traditional agricultural management practices; and (ii) to compare present patterns of vascular plant distributions with information on past land-use derived from a map from 1865.

# Methods

## *Study areas and sampling design*

The study areas included in this thesis are located within the SE part of Norway (Papers I-V), comprising the neighbouring counties Østfold, Vestfold, Akershus, Hedmark and Oppland, in addition to one site in W Norway (Sogn og Fjordane; Paper VI) (Fig. 1).

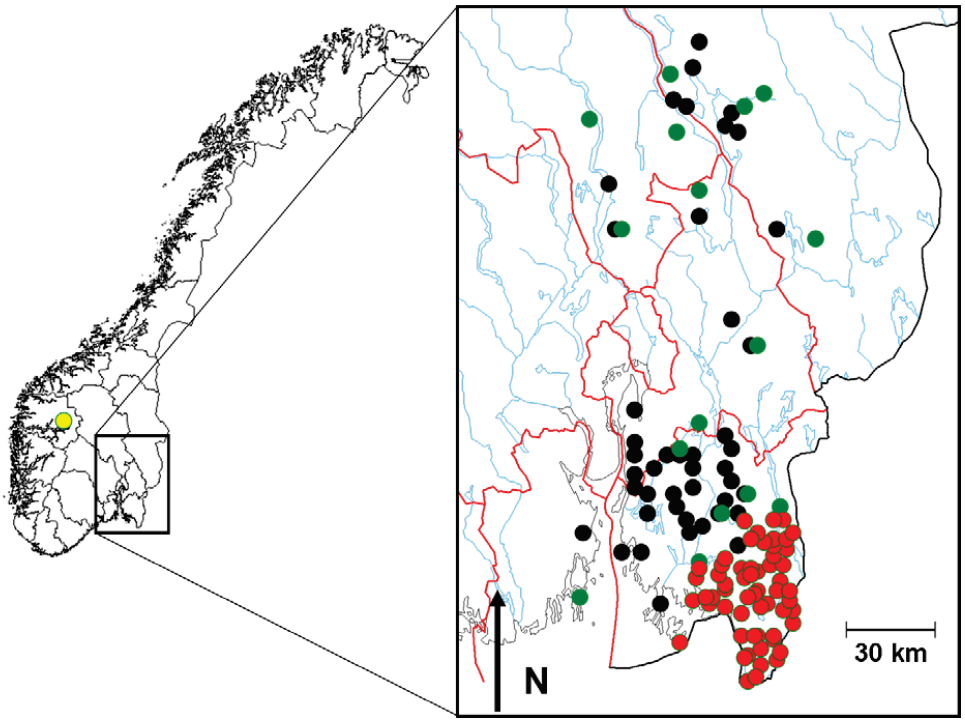


Fig. 1. Map showing all study sites; 2201 patches located within 16 1-km$^2$ plots in green (Papers I and II), 237 *Scorzonera humilis* patches located within 61 1-km$^2$ squares in red (Papers III and V), 63 farm ponds located within 43 sites in black (Paper IV) and the Ornes study area in yellow (Paper VI).

The agricultural landscape of the study area in SE Norway (Papers I-V) is mostly flat and open, surrounded by more or less extensive woodlands. All sites were located in the lowland boreal, boreo-nemoral or southern boreal vegetation zones and situated in the indifferent, markedly or slightly oceanic vegetation section (Moen et al. 1999). The study sites which are described in Paper I-V will be referred to as modern agricultural

landscapes. They more or less comprise patches of larger areas which have been subjected to intensive land-use practices like ploughing, sowing, fertilisation and herbicides during the past decades (Fig. 2).

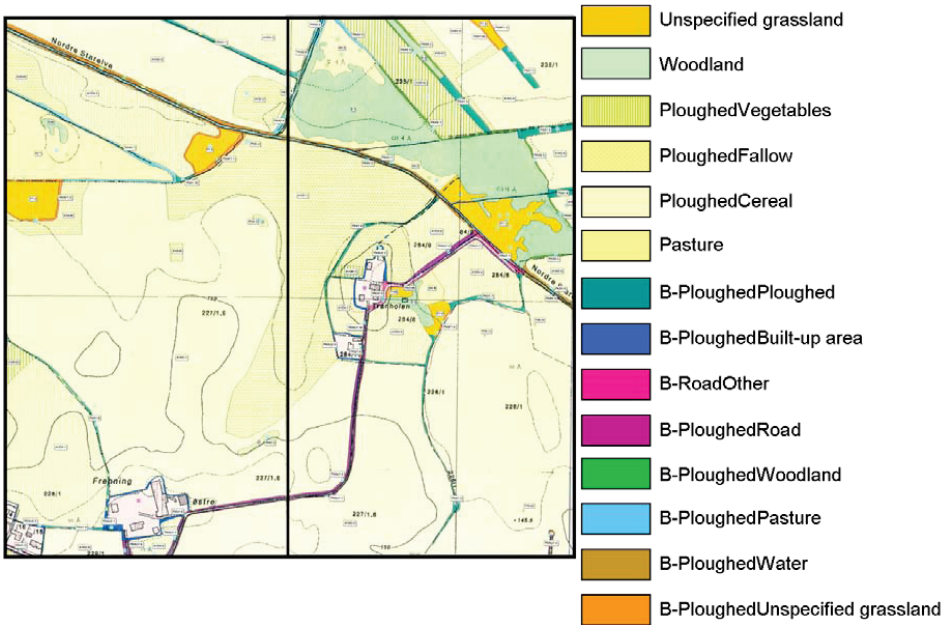


Fig. 2. Distribution of different patch types in the 3Q study area (Papers I and II), exemplified by this 1-km$^2$ square from Stange, Hedmark. See Paper II for detailed information on patch types.

The Ornes study site (Paper VI) is located in Sogn og Fjordane, W Norway, and belongs to the southern boreal and slightly oceanic vegetation zone (Moen et al. 1999). This study area has been created by traditional agricultural management practices and still contains some patches that are mown (mulched) and that have not been cultivated or fertilised by mineral fertiliser. Even though various land-cover and structural changes have occurred during the past centuries, much of the long-established structural pattern is still present, as judged by comparison of the present patch structure with a cadastral map from 1865. This study area will thus be referred to as a fine-scaled, more or less traditional agricultural landscape (Fig. 3).

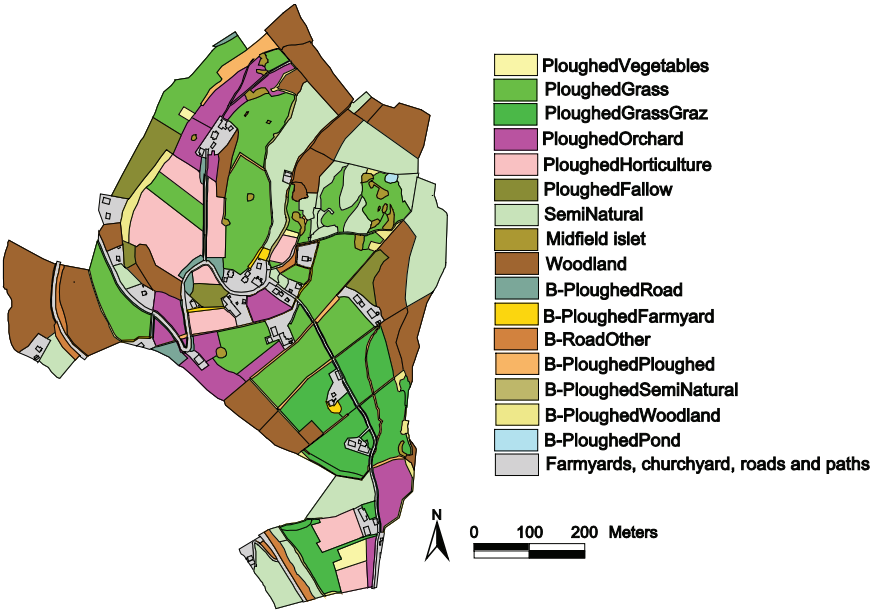


Fig. 3. Distribution of the different patch types in the Ornes study area mapped in 2002 (from Hamre et al., 2007).

Papers I and II are based upon analysis of an extensive data set that includes 2201 patches. These were picked from 16 randomly selected 1-km$^2$ plots implemented in the national monitoring programme '3Q' (Dramstad et al. 2002). The 3Q-plots are distributed across the country in proportion to the area cover of agricultural land and spatial heterogeneity of land-type patches (see Bratli et al. 2006 for more details). A similar procedure based upon 3Q was used to select a total of 63 farm ponds and their adjacent margins included in Paper IV. The *Scorzonera humilis* data set (Papers III and V) comprises 237 patches, all of which have been identified during field work in the six southernmost municipalities of Østfold county. Finally, the Ornes study area included in Paper VI comprises 205 patches which were included in a land consolidation process in 1865 and is therefore also covered by an old cadastral map (scale 1:2 000).

An *a priori* classification of the landscape into patch and boundary types was made for the study sites included in Papers I-II, IV and VI. This categorization was based on a slight modification of the classification system used in the 3Q programme (see Bratli et al. 2006 for more details). Detailed maps showing patch structure were constructed by interpretation of aerial photographs (Papers I-II and IV) or existing economic map (N5, scale 1:5 000; Paper VI) followed by field validation. Patches holding *Scorzonera humilis*

(Papers III and V) were classified according to Naturtyper i Norge ('Nature types in Norway'; NiN; Halvorsen et al. 2009; see Paper III for details about delimitation of these patches). All species were registered present or absent in each patch, except from the Ornes study area where presence of vascular plants was recorded using an ordinal scale (1 = scattered, 2 = widely spread and 3 = dominant).

Relevant environmental variables such as patch geometry, habitat diversity, geographical positioning, various ecological factors including present and historical land-use and human impact variables were recorded for each patch (see individual papers for more details). All variables were zero skewness transformed (Økland et al. 2001) in order to improve homoscedasticity and then brought onto a standard 0–1 scale by ranging.

## *Statistical analyses*

A variety of statistical methods was used in analysing different data sets in this thesis. Patterns of spatial and temporal dynamics in vascular plant species composition and diversity in the agricultural landscapes were examined by ordination methods, habitat specificity analyses and calculation of the semi-natural affiliation index (Papers I-IV and VI). Additionally, habitat suitability modelling was used to predict the distribution of a semi-natural grassland species, *Scorzonera humilis* (Paper V).

Habitat specificity analyses (see Paper I for details) were performed to examine whether rare species tend to co-occur in the same types of habitat patches in the studied landscape (Paper II). These analyses also did identify the land types that hold the most habitat specific flora. Furthermore, the semi-natural grassland affiliation index (SNI) was used to investigate the distribution of semi-natural grassland species among defined landscape patch types (see Paper II for details on SNI). For the 2201 patches data set (Paper II) split-plot GLM and semivariance analyses were used to partition the variation in habitat specificity at different sampling scales along with analysing spatial patterns.

Species compositional patterns and patterns of species compositional change were analysed by DCA (detrended correspondence analysis; Hill and Gauch, 1980) and GNMDS (global non-metric multidimensional scaling; Minchin, 1987) ordination in parallel, and finally, by interpretation of ordination results the most suitable method was chosen (Papers III, IV and VI; see e.g. Paper III for details on ordination methods). Relationships between explanatory variables and patterns in species composition, habitat

specificity and semi-natural affiliation were interpreted by correlation analyses and generalised linear modelling (GLM).

A habitat suitability model (HSM) for the red-listed plant species *Scorzonera humilis*, which is characteristic of extensively managed semi-natural grasslands, was made and eventually evaluated (Paper V). A maximum-likelihood modelling method based upon the maximum entropy principle was used to produce the habitat suitability model for *Scorzonera humilis* within the study area (see Paper V for details on HSM).

All statistical analyses were carried out using R software (R Core Development Team 2009) and/or ArcGIS 9.3 (Anonymous 2008). Ordination analyses were performed using library vegan (Oksanen et al. 2009). The Maxent programme (Phillips et al. 2006) was used for making HSM.

## Summary of results

**The concept of habitat specificity (Paper I)**

In this paper we demonstrate that the original habitat specificity concept of Wagner and Edwards (2001) consists of three independent components: core habitat specificity (uniqueness of the species composition), patch area and patch species richness. Habitat specificity can thus be seen as a family of indices that may include either area or richness components, or none or both, and that may open for use of different types of mean in the calculation of core habitat specificity. Core habitat specificity is a beta diversity measure: the effective number of communities in the landscape, whereas habitat specificity weighted by species number is a gamma diversity measure: the effective number of species that a patch contributes to landscape richness. By comparing 12 different habitat specificity indices by theoretical reasoning and by use of field data our results show that habitat specificity indices are strongly influenced by weights of patch area and patch species richness, in addition to the relative contribution of rare vs. common species (type of mean).

**Habitat specificity in modern agricultural landscapes (Paper II)**

By analysing species data from 2201 patch elements in SE Norwegian modern agricultural landscapes we find that the context used for measuring habitat specificity strongly influences the results. In general the gamma diversity contribution and core habitat specificity calculated from the patch data set were correlated, but this did not hold true for ploughed land patch types. High values for both measures were observed for woodland, pastures and road verges whereas midfield islets and boundary transitional types were ranked low, as opposed to findings in some traditional, extensively managed agricultural landscapes. The semi-natural affiliation index also peaked and dropped in the corresponding patch types. On the other hand results obtained by use of checklist data from the same study area diverged from patch data; ploughed land types were ranked high, whereas woodland was ranked low. Several environmental variables, patch geometry and habitat diversity were significant predictors of habitat specificity. Light and nitrogen were negatively related to habitat specificity whereas moisture, habitat diversity and patch geometry (area, perimeter and shape) influenced habitat specificity positively.

**The present distribution of a rare, semi-natural grassland species in a modern agricultural landscape: exploitation of new habitats or the ghost of meadows past? (Paper III)**

Half of the species observed in patches with *Scorzonera humilis* obtained positive semi-natural affiliation index values, demonstrating that many of the species associated with extensive management and semi-natural grassland habitats are still present within modern agricultural landscape patches. Historical land-use may thus influence species composition for several decades after direct management has come to an end, as was the case for almost all of the investigated patches. One major floristic gradient was identified by ordination methods and that was related to regrowth succession and nutrient status. The short gradient in addition to strong overlap between patch types indicate that the studied patch types, for instance roadsides and semi-natural grasslands, have many species in common. The results indicate a time-lagged response to abandonment, which offers an opportunity for many species to disperse into other patch types in the modern agricultural landscape, if suitable patches exist. Such patches will, however, be impacted by the surrounding landscape matrix.

**Temporal and spatial dynamics of vascular plants in SE Norwegian farm ponds (Paper IV)**

Species compositional changes in farm ponds during a moderately short time interval (six years) were small. We therefore inferred from the results that chance events are important, that the system as such has high degree of unpredictability and that more than six years are needed to detect significant species occurrence trends for a majority of species in this ecosystem. The species composition seems mainly to be determined by idiosyncratic combinations of environmental variables, historical events and random factors, specific for each pond. Once the vegetational community has been established, management practices and environmental factors play less important roles. On the other hand, some significant species compositional changes could be seen in the adjacent pond margins, related to a geographical gradient in addition to eutrophication. This was also in accordance with the observation that an eutrophication has taken place within the six-year study period.

**A fine-scaled habitat suitability model for the red-listed vascular plant *Scorzonera humilis* (Paper V)**

Habitat suitability modelling (HSM) performed by Maxent revealed a high AUC value (0.964) which corresponded to good or excellent model performance. The predictor variables distance to nearest road and cover by arable land contributed most strongly to the Maxent model. There was a significant positive relationship between relative predicted probabilities of occurrence and true probability of presence tested by field evaluation data. Nevertheless, prevalence was low; *Scorzonera humilis* was only observed in 14 of the 528 sample plots in the evaluation data set and the evaluation of the Maxent model by the independent evaluation data set resulted in an AUC value of 0.810. Our results strongly emphasize the importance of validating HSM models by use of independently sampled evaluation data. We conclude that habitat suitability modelling may be a cost-efficient supplement to monitoring and inference about the distribution of rare species.

**Plant species richness, composition and habitat specificity in a Norwegian agricultural landscape (Paper VI)**

The main vegetation gradient in species composition primarily reflected long-term management (land-use) intensity and spanned from intensively managed fields to extensively managed semi-natural grasslands and woodlands. The second ordination axis separated open land from patches not in agricultural use. Both species richness and habitat specificity decreased with increasing management intensity. The two measures of habitat specificity in addition to number of species both peaked in semi-natural grasslands. Strong correlations between core habitat specificity and the gamma diversity contribution of individual patches could be seen, with the exception of vegetable fields. Semi-natural grassland species occurred throughout the landscape, although with some concentration to semi-natural grasslands and woodlands. Comparison of the present patterns of vascular plant distributions with information on past land-use derived from a map from 1865, indicated that species typical of semi-natural grasslands persist long after land-use activities have come to an end and that there is an extinction debt in the vascular flora.

## General discussion and concluding remarks

This study demonstrates the applied use of habitat specificity analysis as a tool for partitioning the landscape species diversity on landscape elements by separating patches with many rare specialist species from patches with the same number of species, all of which are common generalists (Papers I, II and VI).

Habitat specificity, corresponding to the original concept used by Wagner and Edwards (2001), is shown to be a family of diversity indices which are influenced by weights like patch area and species richness, in addition to the contribution of rare versus common species (corresponding to 'type of mean') (Paper I). In Paper I we sort out the different components of which habitat specificity is composed, and relate these components to the standard terminology of biological diversity (α, β and γ). We also demonstrate strong conceptual differences between the two definitions of β-diversity, the multiplicative and the additive, and propose a new term ζ (zeta) diversity for the latter, which is the more recent of these. We argue that patch size cannot be distinguished as an attribute directly related to patch distinctiveness. Removing the area and richness components from habitat specificity indices thus leaves core habitat specificity, which is an index of multiplicative patch β-diversity that represents the number of exclusive species in each patch. This is illustrated by an example of a patch holding some very rare species in addition to some common ones. Such a patch will obtain lower values for core habitat specificity than patches with the same number of rare species but no common ones. This indicates that the former patch is less unique because in addition to its very rare species it shares a larger fraction of species with other patches. We show that, in addition to core habitat specificity, the gamma diversity contribution stand out as a particularly valuable indicator of biodiversity.

The positive relationship between core habitat specificity and both area and species richness reveals that patches with few species mainly hold common species and contribute less to landscape β-diversity than patches with more species. Furthermore, species will be added in sequence of increasing rareness with larger areas. This may be attributed to the generally positive species-area relationship (MacArthur and Wilson, 1967). This relationship is often explained by higher diversity of habitat types in larger areas. On the other hand, habitat specificity based upon the harmonic mean and weighted by species richness (Lu et al., 2007) represents the contribution of each patch to landscape (γ)

diversity. Such total species richness indices give prevalence to patches that hold many species *and* rare species. Most notably patches that hold many rare species are highlighted, i.e. patches that contribute most strongly to landscape γ-diversity.

In discussing the relevance of each index for evaluation of patches in a biodiversity context, we conclude that both of the proposed habitat specificity measures are useful as indicators of biological diversity. Furthermore, we propose that they should be used in parallel to obtain more or less value neutral diversity assessments (Paper II). We also propose that core habitat specificity is an ecologically interpretable measure that specifically addresses patch uniqueness, while habitat specificity weighted by species number combines species richness and species composition in ways relevant for conservation biological assessment. However, the use of context data is of high importance (Paper II). The use of checklist data may for instance alter the outcome of habitat specificity index calculations. Caution is therefore needed in interpretation of habitat specificity results because modern agricultural landscapes contain several land types which are seldom surveyed by botanists, thus being under-represented in checklist (and other museum-type) data. This may be exemplified by ploughed land types such as vegetable and cereal fields which often hold a specific flora compared other patch types in the surrounding landscape. Species typical of such patch types are less documented in herbarium and checklist data and this will thus affect results of species diversity assessments based upon such data.

In general agricultural landscapes hold many red-listed species (Kålås et al., 2006). As discussed above, the use of habitat specificity indices may consequently provide information of relevance to conservation goals at regional levels, in addition to the classification of locally and nationally rare species. Both measures of habitat specificity (based on the 2201 patches data set; Papers I and II) demonstrate that woodland and semi-natural areas stand out as key indicators of species diversity in the modern agricultural SE Norwegian landscape (Paper II). Moreover, woodlands are the only patch type in this data set which obtains a positive semi-natural affiliation value. This implies that the woodlands hold a higher proportion of species typical of semi-natural grasslands with long-term extensive management than all other patch types, including pastures, midfield islets and roadsides (see below) (Paper II). Moreover, these results for semi-natural grasslands and woodlands also, quite surprisingly, hold true for traditionally managed landscape patches (Paper VI). Forests located within the matrix of agricultural landscape areas may be of different origin. They may be secondary forests and some may even still be open and

contain semi-natural species because they may have been kept open due to outfield grazing in the recent past (Bjor and Graffer, 1963). Additionally, road verges, pastures and unspecified grassland areas show relatively high values for both habitat specificity indices and also possess locally and nationally rare species (Papers II and VI). This demonstrates that the presence of natural and semi-natural habitats increases the richness of habitat specialists in many types of patches within agricultural landscapes.

On the other hand, the low index values of semi-natural grassland affiliation for many patch types demonstrate that pastures, road verges and unspecified grassland areas on average contain relatively equal proportions of species affiliated with traditional semi-natural management practices and intensively managed land (Papers II and VI). This also holds true for midfield islets patches included in the 2201 patches data set (Paper II). Such results contrast the common view that agricultural landscape margins are often associated with high diversity (Burel, 1996). Semi-natural species are in general light-demanding, they prefer unfertilised soil and depend on management intensity and management practices like mowing (Norderhaug et al., 2000). Most marginal areas like field boundaries and midfield islet today are no longer regularly mown, and they may be exposed to fertiliser and herbicides. Several such patches have a species composition that resembles that of intensively used agricultural land (Bratli et al., 2006); they are species poor, hold few unique species and are of low value from a biodiversity conservation point of view.

In contrast, all patch types included in the *Scorzonera humilis* study (Paper III) obtained positive values for the semi-natural grassland affiliation index (SNI). The overall mean SNI value of 0.76 should nonetheless be considered high compared to the patches included in the traditionally managed landscape of W Norway. In the traditional landscape of W Norway a mean SNI of 0.34 was found both for semi-natural grassland and forest patches, whereas roadsides obtained SNI values between –0.25 and 0.25 (Paper VI). Moreover, as discussed above, only negative SNI values were obtained for all patch types included in Paper II except for forests (SNI = 0.29). These apparently paradoxical results may have several explanations, among others differences (natural or historical) among data sets and study areas. The study area in which *S. humilis* patches are located (Paper III) involves a rather fine-scaled agricultural landscape thereby including numerous borderlines between a variety of patch types compared to more intensively used and large-scale landscapes of Paper II. The findings may also be attributed by the negative relationship between habitat specificity and nitrogen availability (Paper II). Such relationship can be explained by effects of intensive agricultural activity on the flora of patches (within and)

adjacent to arable land, due to fertilisation (Willi et al., 2005). Numbers of ruderal, nitrophilous species such as competitive grasses and weedy perennials increase, eventually on the cost of overall biodiversity. This may explain why the 3Q study area of mainly modern, mostly large-scale agricultural landscapes (Paper II) holds a more trivial species composition in addition to many ruderals being favoured by fertility and disturbance. Furthermore, this also strongly underlines that the studied landscape of SE Norway (Paper II) is a modern agricultural landscape in the sense that traditionally managed patches hardly occur. Additionally, the *S. humilis* patches are selectively picked; in general they are all being characterised by semi-natural species like *S. humilis* and thus do not comprise a representative diversity of patch types within one particular agricultural landscape area. This may also contribute to the high SNI.

Nevertheless, there are apparent discrepancies between the rather extensively used patches in western Norway (Paper VI) compared to the more intensively used SE Norwegian *S. humilis* patches (Paper III), although both of them are rather fine-scaled. The findings of relatively high semi-natural affiliation index values emphasise that modern agricultural landscapes may also stand out as important from a diversity point of view. They may comprise a multiplicity of land patch types holding a number of exclusive species. However, this needs not necessarily to be universal for all such types of landscapes but will depend on historical factors and the rate of land-use change.

This study (Paper VI) also underlines the importance of old cadastral maps as an important source of landscape historical information. Such maps provide information on spatiotemporal landscape changes and it is demonstrated that present diversity patterns reflect past land-use and management. The high abundance of semi-natural affiliated species both in abandoned semi-natural grassland areas and forests today (Paper III) supports the view that due to slow response to altered environmental conditions, many species of semi-natural grasslands remain in such habitats for a long time after cessation of management (Tilman et al. 1994; Cousins et al., 2007). However, even if such species may still be present within deteriorating habitats and thereby subjected to very small long-term abundance changes, they will eventually go extinct if the habitat is abandoned and the habitat quality is reduced. Subsequently, these patches will be increasingly impacted by the surrounding landscape matrix. A landscape demonstrating large variation in habitats is thus desirable, comprising both small and large patch elements. Semi-open forests, roadsides and remnant margin areas may act as refugia as well as source populations in larger meta-communities for semi-natural affiliated species also belonging to the traditional

agricultural landscape which is now rapidly disappearing (Milberg and Persson, 1994). However, it is clear that appropriate management practices in a landscape perspective will therefore be highly important for maintenance of the characteristic vascular plant flora of semi-natural grasslands (Paper III).

As discussed above, semi-natural grasslands are biodiversity hotspots in the agricultural landscape, also comprising rare species (Losvik, 2007). To be able to detect, and thereby monitor and manage, rare or red-listed species habitat suitability modelling (HSM) may be used (Paper V). Rapid recent developments in HSM methodology demonstrate considerable opportunities for improvements in this field (see, e.g. Araújo and Guisan, 2006). For instance, AUC has long been referred to as the standard measure of HSM performance and is considered a good overall indicator of modelling success (Elith et al., 2006). In this study we propose that other properties of the receiver operating characteristic (ROC) curve should also be reported; like maximised sum threshold (MST), minimum difference threshold (MDT), sensitivity and specificity values corresponding to different relative predicted probabilities of occurrence thresholds. The present study (Paper V) confirms results of previous studies (e.g. Elith et al., 2006; Guisan et al. 2006) that Maxent produces meaningful HSM models. High predictive power may be due to fine-scaled resolution of data (25 m linear scale), species occurrence as well as predictor variable data. Additionally, to avoid uncertainties in the spatial prediction model of *Scorzonera humilis* this work exclusively makes use of exact, geo-referenced presence observations confirmed by field work. Furthermore, we demonstrate that independent evaluation data set(s) should be used to address the reliability of HSM models. Such data should preferably be collected in the same study area. By involving independent data, the prevalence of the species (frequency in the study area) can be estimated. Additionally, the relative predicted probability of occurrence (RPPO) values produced by the HSM model can be re-interpreted in terms of estimated true probabilities of presence (TPP) in all pixels within the study area. Such quantifications add strongly to the value of HSM models for rare species making them easier to detect and thereby monitor.

The study of farm ponds gives important insight into the dynamics of hydro- and helophytes and terrestrial species associated with such ecosystems in modern agricultural landscapes (Paper IV). In general, ponds and other habitat patches in agricultural landscapes are heterogeneous and characterised by rather fast land-use changes. Our results, though, point to farm ponds as being less dynamic that expected. Despite a shift towards higher degree of eutrophication, few distinct trends in temporal development of

the heterogeneous macrophyte communities could be seen. However, the patterns were somewhat different for hydro- and helophytes in ponds and terrestrial species in pond margins. The slow change in aquatic species composition implies that macrophyte communities are strongly contingent on each pond's initial vegetation during the first years of establishment. Nevertheless, changes may take place if exceeding a threshold of nutrient content. Besides, pond ecosystems are often exposed to disturbances which are likely to alter species compositions and create open sites for re-colonisation (Grubb, 1977). In addition to current environmental conditions, the plant species composition of ponds will thereby be determined by historical, often pond-specific events (Jeffries, 2008). The lack of strong temporal patterns within the ponds supports our view of chance and unpredictability as important factors (Edvardsen and Økland, 2006b). Moreover, ponds function as environmental islands in the agricultural landscape (Jennersten et al., 1992). In addition to randomness being a structural factor, such islands represent distinct communities which show independent development; a process which is also weakening temporal trends.

I conclude that knowledge on spatial and temporal patterns of species distribution, correspondingly involving diversity and composition, in agricultural landscapes, is highly important. Several of the studies that make up this thesis show that key elements of species diversity can be identified by habitat specificity measures which can thereby create a basis for both conservation and value-neutral practices in modern as well as traditional agricultural landscapes. With comprehensive field surveys of species composition and detailed information on patch configuration and other environmental variables, e.g. obtained from old cadastral maps, diversity patterns and habitat suitability models can be derived. The thesis shows that results obtained by these tools may provide valuable insight in species distributional patterns of whole landscapes, including past, present and potential future perspectives.

## Acknowledgements

I would like to thank Inger Auestad, Marianne Evju and Rune Halvorsen for helpful comments on this thesis.

## References


Anonymous. 2002. Jordbruksteljing 1999. NOS C 670. Statistics Norway, Oslo.

Anonymous. 2008. ArcGIS version 9.3. ESRI Inc. Redlands, California.

Anonymous. 2009. The World Factbook 2009. Central Intelligence Agency, Washington DC. https://www.cia.gov/library/publications/the-world-factbook/index.html

Araújo, M.B., Guisan, A. 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr., 33: 1677-1688.

Berge, G., Hestmark, G. 1997. Seed banks of roadsides, stream verges and fields. Ann. Bot. Fennici 34: 77-90.

Bjor, K., Graffer, H. 1963. Beiteundersøkelser på skogsmark. Forsk. Fors. Landbr. 14: 121-365.

Bratli, H., Økland, T., Økland, R.H., Dramstad, W.E., Elven, R., Engan, G., Fjellstad, W., Heegaard, E., Pedersen, O., Solstad, H. 2006. Patterns of variation in vascular plant species richness and composition in SE Norwegian agricultural landscapes. Agricult. Ecosyst. Environm. 114: 270-286.

Brys, R., Jacquemyn, H., Endels, P., de Blust, G., Hermy, M. 2004. The effects of grassland management on plant performance and demography in the perennial herb *Primula veris*. J. Appl. Ecol. 41: 1080-1091.

Burel, F. 1996. Hedgerows and their role in agricultural landscapes. Crit. Rev. Plant Sci. 15: 169-190.

Cousins, S.A.O., Ohlson, H., Eriksson, O. 2007. Effects of historical and present fragmentation on plant species diversity in semi-natural grasslands in Swedish rural landscapes. Landsc. Ecol. 22: 723-730.

Cousins, S.A.O., Lindborg, R. 2008. Remnant grassland habitats as source communities for plant diversification in agricultural landscapes. Biol. Conserv. 141: 233-240.

Cousins, S.A.O. 2009. Extinction debt in fragmented grasslands: paid or not? J. Veg. Sci. 20: 3-7.

Dramstad, W.E., Fjellstad, W.J., Strand, G.-H., Mathiesen, H.F., Engan, G., Stokland, J., 2002. Development and implementation of the Norwegian monitoring programme for agricultural landscapes. J. environm. Mgmt. 64: 49-63.

Edvardsen, A., Økland, R.H. 2006a. Variation in plant species richness in and adjacent to 64 ponds in SE Norwegian agricultural landscapes. Aquat. Bot. 85: 79-91.

Edvardsen, A., Økland, R.H. 2006b. Variation in plant species composition in and adjacent to 64 ponds in SE Norwegian agricultural landscapes. Aquat. Bot. 85: 92-102.

Edvardsen, A., Halvorsen, R., Norderhaug, A., Pedersen, O., Rydgren, K. Habitat specificity of patches in modern agricultural landscapes. Landsc. Ecol., in press. DOI 10.1007/s10980-010-9481-2

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Townsend Peterson, A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Farina, A. 1998. Principles and methods in landscape ecology. Chapman & Hall, London.

Grubb, P.J. 1977. The maintenance of species-richness in plant communities: the importance of the regeneration niche. Biol. Rev. 52: 107-145.

Guisan, A., Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Modelling 135: 147-186.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E. 2006. Using niche-based models to improve the sampling of rare species. Conserv. Biol. 20: 501-511.

Halvorsen, R., Andersen, T., Blom, H.H., Elvebakk, A., Elven, R., Erikstad, L., Gaarder, G., Moen, A., Mortensen, P.B., Norderhaug, A., Nygaard, K., Thorsnes, T., Ødegaard, F. 2009. Ny norsk naturtypeinndeling version 1.0 http://www.naturtyper.artsdatabanken.no

Hamre, L.N., Domaas, S.T., Austad, I., Rydgren, K. 2007. Land-cover and structural changes in a western Norwegian cultural landscape since 1865, based on an old cadastral map and a field survey. Landsc. Ecol. 22: 1563-1574.

Hill, M.O., Gauch, H.G. Jr. 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42: 47-58.

Jeffries, M. 2008. The spatial and temporal heterogeneity of macrophyte communities in thirty small, temporary ponds over a period of ten years. Ecography 31: 765-775.

Jennersten, O., Loman, J., Møller, A.P., Robertson, J., Widen, B. 1992. Conservation biology in agricultural habitat islands. In: Hansson, L. (Ed.). Conservation biology in agricultural habitat islands. Elsevier, Barkham, pp. 395-423.

Klimek, S., Kemmermann, A.R.G., Hofmann, M., Isselstein, J. 2007. Plant species richness and composition in managed grasslands: the relative importance of field management and environmental factors. Biol. Conserv. 134: 559-570.

Kålås, J.A., Viken, Å., Bakken, T. (Eds.). 2006. Norsk rødliste 2006 - 2006 Norwegian red list. Artsdatabanken, Trondheim.

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76: 5-13.

Liira, J., Schmidt, T., Aavik, T., Arens, P., Augenstein, I., Bailey, D., Billeter, R., Bukáček, R., Burel, F., De Blust, G., De Cock, R., Dirksen, J., Edwards, P., Hamerský, R., Herzog, F., Klotz, S., Kühn, I., Le Coeur, D., Miklová, P., Roubalova, M., Schweiger, O., Smulders, M.J.M., van Wingerden, W.K.R.E., Bugter, R., Zobel, M. 2008. Plant functional group composition and large-scale species richness in European agricultural landscapes. J. Veg. Sci. 19: 3-14.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson., J., Grime, J.P., Hector, A. Hooper, D.U., Huston, M.A., Raffaelli, D., Schmid, B., Tilman, D., Wardle, D.A. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenge. Science 294: 804-808.

Losvik, M. 2007. Regional species pools of hay meadows: a case study. Appl. Veg. Sci. 10: 239-248.

Lu, H.-P., Wagner, H.H., Chen, X.Y. 2007. A contribution diversity approach to evaluate species diversity. Basic Appl. Ecol. 8: 1-12.

Luoto, M., Rekolainen, S., Aakkula, J., Pykälä, J. 2003. Loss of plant species richness and habitat connectivity in grasslands associated with agricultural change in Finland. Ambio 32: 447-452.

MacArthur, R., Recher, H., Cody, M. 1966. On the relation between habitat selection and species diversity. Am. Nat. 100: 319-332.

MacArthur, R.H., Wilson, E.O. 1967. The theory of island biogeography. Princeton Univ. Press, Princeton.

Magurran, A.E. 1988. Ecological diversity and its measurement. Croom Helm, London.

Milberg, P., Persson, T.S. 1994. Soil seed bank and species recruitment in road verge grassland vegetation. Ann. Bot. Fennici 31: 155-162.

Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69: 89-107.

Moen, A., Lillethun, A., Odland, A. 1999. National atlas of Norway: vegetation. Norwegian Mapping Authority, Hønefoss.

Norderhaug, A., Austad, I, Hauge, L, Kvamme, M. 1999. Skjøtselsboka for kulturlandskap og gamle norske kulturmarker. Valdres Trykkeri, Fagernes.

Norderhaug, A., Ihse, M., Pedersen, O. 2000. Biotope patterns and abundance of meadow plant species in a Norwegian rural landscape. Landsc. Ecol. 15: 201-218.

Noss, R.F. 1990. Indicators for monitoring biodiversity: a hierarchical approach. Conserv. Biol. 4: 355-364.

Økland, R.H. 1990. Vegetation ecology: theory, methods and applications with reference to Fennoscandia. Sommerfeltia Suppl. 1: 1-233.

Økland, R.H., Økland, T., Rydgren, K. 2001.Vegetation-environment relationships of boreal spruce swamp forests in Østmarka Nature Reserve, SE Norway. Sommerfeltia 29: 1-190.

Økland, R.H., Bratli, H., Dramstad, W.E., Edvardsen, A., Engan, G., Fjellstad, W., Heegaard, E., Pedersen, O., Solstad, H. 2006. Scale-dependent importance of environment, land use and landscape structure for species richness and composition of SE Norwegian modern agricultural landscapes. Landsc. Ecol. 21: 969-987.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B. 2007. Package 'vegan' Version 1.9-13. Univ. of Oulu, Oulu.

Peet, R.K. 1974. The measurement of species diversity. A. Rev. Ecol. Syst. 5: 285-307.

Phillips, S.J., Anderson, R.P., Schapire, R.E. 2006. Maximum entropy modelling of species geographic distributions. Ecol. Modelling 190: 231-259.

Pielou, E.C. 1975. Ecological diversity. Wiley, New York.

Poschlod, P., Bakker, J.P., Kahmen, S. 2005. Changing land use and its impact on biodiversity. Basic. Appl. Ecol. 6: 93-98.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

Robinson, R.A., Sutherland, W.J. 2002. Post-war changes in arable farming and biodiversity in Great Britain. J. Appl. Ecol. 39: 157-176.

Shannon, C.E., Weaver, W. 1949. The mathematical theory of communication. Univ. of Illinois Press, Urbana.

Simpson, E.H. 1949. Measurement of diversity. Nature 163: 688.

Smart, S.M., Bunce, R.G.H., Firbank, L.G., Coward, P. 2002. Do field boundaries act as refugia for grassland plant species diversity in intensively managed agricultural landscapes in Britain? Agricult. Ecosyst. Environm. 91: 73-87.

Stohlgren, T.J., Coughenour, M.B., Chong, G.W., Binkley, D., Kalkhan, M.A., Schell, L.D., Buckley, D.J., Berry, J.K. 1997. Landscape analysis of plant diversity. Landsc. Ecol. 12: 155-170.

Strand, G.-H., Bekkhus, R. 2008. Markslagsstatistikk. Dyrka og dyrkbart areal. Ressursoversikt fra Skog og landskap 03/08: 1-33.

Strand, G.-H., Eriksen, R. 2008. Markslag og skogstatistikk. Jordbrukets kulturlandskap. Ressursoversikt fra Skog og landskap 01/08: 1-47.

Strijker, D. 2005. Marginal lands in Europe –causes of decline. Basic Appl. Ecol. 6: 99-106.

Tilman, D., May, R.M., Lehman, C.L., Nowak, M.A. 1994. Habitat destruction and the extinction debt. Nature 371: 65-66.

Van Dijk, G. 1991. The status of semi-natural grasslands in Europe. In: Goriup, P.D., Batten, L.A., Norton, J.A. (Eds.). The conservation of lowland dry grassland birds in Europe. Joint Nature Conservation Committee, Peterborough, pp. 15-36.

Vandvik, V., Birks, H.J.B. 2002. Partitioning floristic variance in Norwegian upland grasslands into within-site and between-site components: are the patterns determined by environment or by land-use? Plant Ecol. 162: 233-245.

Wagner, H.H., Edwards, P.J. 2001. Quantifying habitat specificity to assess the contribution of a patch to species richness at a landscape scale. Landsc. Ecol. 16: 121-131.

Waldhardt, R. 2003. Biodiversity and landscape – summary, conclusions and perspectives. Agricult. Ecosyst. Environm. 98: 305-309.

Whittaker, R.H. 1960. Vegetation of the Siskiyou mountains Oregon and California. Ecol. Monogr. 30: 279-338.

Whittaker, R.H. 1967. Gradient analysis of vegetation. Biol. Rev. Camb. Phil. Soc. 42: 207-264.

Whittaker, R.H. 1972. Evolution and measurement of species diversity. Taxonomy 21: 213-251.

Willi, J.C., Mountford, J.O., Sparks, T.H. 2005. The modification of ancient woodland ground flora at arable edges. Biodivers. Conserv. 14: 3215-3233.

# I

# II

III

# IV

V

**A fine-scaled habitat suitability model for the red-listed vascular plant *Scorzonera humilis***

Anette Edvardsen[1], Vegar Bakkestuen[1] and Rune Halvorsen[1]

[1]Dept of Research and Collections, Natural History Museum, University of Oslo

PO Box 1172 Blindern, NO-0318 Oslo, Norway.

e-mail (corresponding author): anette.edvardsen@nhm.uio.no

phone: +47 22 85 18 16/fax: +47 22 85 18 35

## Abstract

We aim at exploring the applied use of niche modelling as a tool for making spatial predictions of occurrences of red-listed vascular plant species on a small scale (25 m × 25 m resolution) using accurate georeferenced data for all records, also discussing the use of independent evaluation data for such models. Maxent was performed to make habitat suitability models for a red-listed plant species being characteristic of extensively managed semi-natural grasslands; *Scorzonera humilis*. Three sets of model performance were assessed; training data available from data-splitting in Maxent, independent evaluation data collected during field work in addition to an extended evaluation data set also from the study area. AUC values corresponded to good or excellent model performance. There was a significant positive relationship between relative predicted probabilities of occurrence and true probability of presence, nevertheless, prevalence was low. However, AUC is strongly data-dependent and other properties of the receiver operating characteristic (ROC) curve like maximised sum threshold (MST), minimum difference threshold (MDT), sensitivity and specificity values corresponding to different relative predicted probabilities of occurrence thresholds are thus reported. The present study demonstrates that habitat suitability modelling (HSM) can be seen as a cost-efficient supplement to monitoring and inference about rare species distribution. However, evaluation of models is highly important and independent evaluation data add strongly to HSM performance for rare species.

## Introduction

Loss of biological diversity at ever increasing rates is still an ongoing process all over the world, mainly due to increasing human impact (Baillie et al. 2004). Redlisting, based on evaluation of the risk of species becoming locally or globally extinct, has become an important tool in national as well as international biodiversity conservation (Rodrigues et al. 2006). The nature of being rare implies that red-listed species often have low prevalence (frequency of occurrence) within their distribution area, regardless of the extent of this area being restricted or larger (Guisan et al. 2006a). Obtaining representative information of status and trends for populations of rare species therefore tends to be time consuming and cost-inefficient because a large proportion of potential sampling sites fails to contain the target species. Cost-efficient supplementary alternatives to inference about rare species distribution and population trends are therefore needed (Guisan et al. 2006a).

Models by which the known geo-referenced occurrences of a species is used to characterize the species' distribution along the main niche gradients, which is in turn used to predict the species' potential distribution in geographical space, have become a promising tool for many applied aspects of ecology (Elith et al. 2006, Wisz et al. 2008). Such models are alternatively referred to as habitat suitability models (HSM; Hirzel et al. 2006), habitat distribution models (HDM; Guisan and Zimmermann 2000), species distribution models (SDM; Araújo and Guisan 2006), and ecological niche models (ENM; Peterson 2003). Because the output of such models is relative habitat suitability indices (Hirzel and Le Lay 2008), the term HSM will be used throughout this paper. HSM has proved particularly promising for studies of rare plant dynamics because good spatial predictions may increase sampling efficiency and assisting detection of new occurrences of rare species (Guisan et al. 2006ab, Marage et al. 2008, Parolo et al. 2008). Rare species are, however, among the most difficult to model by HSM methods (Edwards et al. 2005).

During the last century land-use changes have completely altered traditional agricultural landscapes all over Europe. Formerly intensively used areas have become even more intensively used, while marginal areas have been abandoned (Haines-Young et al. 2003). This ongoing polarisation in the landscape has resulted in decline for vascular plant species associated with traditionally managed, semi-natural grasslands (Luoto et al. 2003), to the extent that more than 25 % of species on the 2006 Norwegian Red List are considered as threatened by abandonment and 20 % are threatened by intensification of agricultural

practices (Kålås et al. 2006). Information about population trends, e.g. by HSM methods, are therefore urgently needed for species associated with agricultural land, for most of which the present-day distribution is dependent on management (Cousins and Eriksson 2002). Spatial prediction is, however, particularly difficult for rare species that are dependent on management, because management is a complex variable with a strong temporal dimension, for which proper quantitative information is difficult to obtain. Nevertheless, recent HSM studies of management-dependent vascular plants by Marage et al. (2008) and Parolo et al. (2008) indicate that HSM may be a valuable tool for improving our knowledge of distributional trends for such plant species.

Like all other models, niche models may vary from poor and misleading to excellent and informative (Elith et al. 2006, Wisz et al. 2008). Before applied, niche models need to be evaluated, preferably by use of independent test data (Austin 2007). For rare species present in few sites obtaining such data is, however, more difficult as well as and more important than for other models (Marage et al. 2008). The high importance of evaluation on independent data for such models is due to stronger differences between performance indices based upon resubstitution of training data and indices based upon independent evaluation data (cf. Marage et al. 2008) and the lower reliability of models trained on sparse than on comprehensive presence data (Wisz et al. 2008).

The aim of this study is to explore the applied use of habitat suitability modelling as a tool for making spatial predictions of occurrences of a red-listed vascular plant species, *Scorzonera humilis*, which is associated with semi-natural grasslands in Norway. We use field-validated presence-only (PO) data for our modelling, tightly integrated within a framework of geographical information systems (GIS) as recommended by Bakkestuen (2009). Models were trained on fine-scaled resolution environmental and species information (25 m linear scale) and evaluated by use of independent data obtained by a field-intensive sampling procedure (observations of presence or absence of the modelled species at more than 500 georeferenced points). A main purpose of the study is to discuss the reliability of HSM models for rare species.

## Material and methods

### The species

The herbaceous perennial *Scorzonera humilis* L. (Asteraceae) is a characteristic species of extensively managed semi-natural grasslands (Halvorsen et al. 2009), however, it can typically also be found in roadsides (A. Edvardsen et al., unpublished data). Its distribution in Norway is restricted to the south-eastern part of the country (Lid and Lid 2005), and the species is listed as vulnerable (VU) in the 2006 Norwegian Red List (Kålås et al. 2006). Due to the polarisation in agricultural landscapes during the past decades, a considerable fraction of the area previously covered by semi-natural grasslands has undergone substantial land-use change. Consequently *S. humilis* is at present threatened by agricultural intensivation like ploughing and/or exposition to fertiliser, as well as cessation of management and substantial regrowth. However, due to a slow response to altered environmental settings, such as regrowth, many species (including *S. humilis*) remain in such habitats for many decades (A. Edvardsen et al, unpublished data.).

**Study area**

Less than 10 *S. humilis* localities in total, of which only one was discovered after year 1900, are known from Norway outside Østfold county (www.gbif.no). The study area was therefore restricted to the six south-easternmost municipalities of Østfold county, SE Norway, covering 2305 $km^2$ (Fig. 1) The study area is situated mainly in the boreo-nemoral vegetation zone (with few pendants into the south boreal vegetation zone) and the weakly oceanic vegetation sector (Moen 1998) and is characterised by modern agricultural practices..

**Collection of training data for modelling *S. humilis* presence**

We used all known geo-referenced observations (herbarium and check-list data available from the Global Biodiversity Information Facility – Norway data portal, www.gbif.no, extracted medio May 2008) as basis for identification of *S. humilis* presence sites. All of these sites, as well as nearby sites that were judged as suitable for the species and locations communicated by amateur botanists, were visited during the summer of 2008. All populations and individuals of *S. humilis* that belonged to the same patch type (see below) *and* that were not separated from the nearest neighbouring plant by more than 20 m were considered as one patch. Consequently *S. humilis* populations and individuals that were separated by more than 20 m were treated as separate patches. In total, *S. humilis* was found in 237 patches (Fig. 1).

The patches varied in area from 10 $m^2$ to 2628 $m^2$. All patches were georeferenced by standard GPS equipment (accuracy ± 7 m). Species presences, recorded in the field as vectors – points (single observations), lines (along roadsides or narrow field margins) or polygons (larger observational units) – were entered into ArcGIS 9.3 (Anonymous 2008). GIS was used to transform presences in vector format into raster grid cells (pixels) of 25 m × 25 m.

Presences of *S. humilis* in the 237 patches will be referred to as the *training data* set.

**Choice and recording of environmental predictors**

For our modelling of the relationship between *Scorzonera humilis* and the environment, we used a set of eight variables that were considered the best possible proxies for important niche factors that possible to obtain with mapping units (pixels) of 25 m × 25 m (Table 1).
The extent of the study area, with small variation in climatic conditions and, hence, high uncertainty of climatic information interpolated from a sparse net of meteorological stations, made us restrict ourselves to altitude as a proxy for temperature-dependent regional variation. Within this study area, temperature decreases while precipitation (and climatic humidity) increases with increasing altitude (Økland 1989). Altitude for each pixel was derived from the 25 m digital elevation model (DEM) from the National Mapping Authorities in Norway. The DEM is originally interpolated from contour lines of 20 m resolution.

Local environmental factors of general importance for plants, such as soil depth, wetness, and warmth supply (e.g. Halvorsen et al. 2009), were represented by five variables. Solar irradiance was estimated at equinox, winter minimum and summer maximum by derivation in ArcGIS based on the DEM. The topographic wetness index (TWI; Moore et al. 1991) was calculated for each pixel to represent the spatial variability of soil moisture. TWI, the natural logarithm of the ratio of the catchment area above the pixel ($m^2 \cdot m^{-1}$) and tan of the terrain slope, is a commonly used surrogate variable for soil moisture.

Ideally, our set of predictors should include (proxies for) all important local basic gradients, such as substrate base cation richness and long-term land use, as well as state variables such as current land-use and re-growth successional state of abandoned farmland. However, area-covering variables that represent these factors were not available. Instead, we used digital land-cover information (*Digitalt markslagskart*, DMK; Bjørdal 2007), available as vector format maps to scale 1:5 000, rasterized to pixel size. Because *S. humilis* is known to be associated with semi-natural grassland (which is not recorded as a land-cover class in DMK), and because the species currently spreads into road verges (A. Edvardsen et al.,

unpublished data.), we used GIS to construct two additional variables from DMK data: distance to roads and distance to arable land.

**Modelling technique**

We used Maxent version 3.1.0, default settings (Phillips et al. 2006, Phillips and Dudík 2008), a maximum-likelihood modelling method based upon the maximum entropy principle (Jaynes 1957), to produce a habitat suitability model (HSM) for *Scorzonera humilis* in the study area. The training data set of *S. humilis* presence observations and the eight environmental predictors were used as model input. Maxent was chosen because it has proven one of the most reliable HSM methods currently available, also compared with methods based upon presence-absence (PA) data (Elith et al. 2006, Guisan et al. 2007, Elith and Graham 2009, Phillips et al. 2009).

The raw Maxent output (relative probabilities $p_i$ that sum to 1 over all pixels in the study area) were converted into a logistic representation as recommended by Phillips and Dudík (2008). Because Maxent is based upon PO data the output represents *relative* suitability values (Phillips et al. 2006) and is referred to as relative predicted probabilities of occurrence (RPPO). Calibration of PO-derived suitability (e.g. RPPO) values to a true-probability scale requires information on the prevalence (frequency in the study area, given the spatial resolution of the grid; Hirzel et al. 2006) of the species (Elith and Graham 2009), which with PO-data only is not known.

**Evaluation and use of the model**

We used properties of the receiver operating characteristic (ROC) curve to evaluate the model (Pearce and Ferrier 2000). With PO data, AUC is calculated from a confusion matrix of presences and random background points (pseudo-absences) and measures the probability that the model will assign a higher habitat suitability value to a randomly selected presence point than to a randomly selected point. With PA data, AUC measures the probability that the model will assign a higher habitat suitability value to a randomly selected presence point than to a randomly selected *absence* point.

AUC is a standard measure of HSM performance which is considered a good overall indicator of modelling success (e.g. Elith et al. 2006, Wisz et al. 2008), although AUC has been criticised among others for dependence on species properties such as extent of

occurrence and prevalence and for taking into account regions of the ROC space without interest for the use of models (Lobo et al. 2008). Accordingly, we also determined threshold ROC values that corresponded to two other statistics derived from the ROC curve (see Jiménez-Valverde and Lobo 2007); the minimum difference threshold (MDT) and the maximised sum threshold (MST), reflecting the ROC values that minimised and maximised differences between sensitivity and specificity, respectively.

Three sets of model performance statistics were obtained. The first set was obtained from the set of 30 % of presence observations in the training data set, set aside by Maxent for model evaluation (Phillips et al. 2006). Evaluation by data-splitting is confounded by sampling bias and other kinds of bias in the data set (Edwards et al. 2006). Testing of HSM models on independently sampled test data is therefore strongly recommended (Austin 2007) but rarely carried out because such data are difficult to obtain, particularly for rare species (Engler et al. 2004, Guisan et al. 2006a, Parolo et al. 2008; but see Marage et al. 2008).

In order to address the reliability of HSM models for rare species, we put considerable effort into collection of an independent *evaluation data* set from the study area. To enable evaluation of the ability of the Maxent model to predict occurrence of *S. humilis* over the entire span of occurrence probabilities, we built the evaluation data set by randomised stratification in ArcGIS, as recommended by Guisan et al. (2006b). We iteratively selected 60 transects, each with nine 25 m × 25 m sample plots at regular 100 m intervals. This distance between sample plots in this evaluation data set was larger than the largest dimension of the largest *S. humilis* patch known from the study area, thus reducing spatial autocorrelation in the evaluation data. Selection of transects started by picking one pixel at random from each of ten decile classes into which the RPPO scale was divided. The ten pixels thereby selected were taken as the central sample plots of one transect each. The frequency distribution of RPPO values of the 90 sample plots on the 10 transects was then inspected. The iteration proceeded by picking a new transect centerpixel at random (but no closer than 1 km to any sample plot previously picked) from the decile class that was most poorly represented in the set of sample plots, followed by updating the frequency distribution. The process was repeated until 60 transects had been obtained. The total number of sample plots was 528 because twelve inaccessible sample plots had to be discarded. Each sample plot was checked for true presence or absence of *S. humilis* during field work in the summer of 2009.

During collection of the evaluation data set, 87 additional observations of *S. humilis* were recorded, mostly along the transects. These records were georeferenced and used as an *extended evaluation data* set (of PO observations). For each pixel in this set, the relative

predicted probabilities of occurrence (RPPO) were obtained by GIS analysis of Maxent output.

AUC values for the set-aside presence data were calculated by Maxent. For the evaluation data set, we calculated AUC, sensitivity, specificity, MST (maximised sum threshold) and MDT (minimised difference thereshold). The evaluation data set was also used to model true probability of presence (TPP) as a function of RPPO values (produced by Maxent). This was done in two ways. Firstly, we used the records of presence or absence of *S. humilis* to calculate the observed frequency of occurrence (OFO) for each decile RPPO class. Secondly, we used generalised linear models (GLM; Venables and Ripley 2002) with logit link and binomial errors (logistic regression) to model TPP as a function of RPPO (see Pearce and Ferrier 2000), under the assumption that the detectability of the target species is 1, i.e. that the species was detected in all pixels in which it was present. The model logit(p^) = f(RPPO) was used to calculate estimates (p^) and 95% confidence intervals for TPP, for all pixels, by the inverse logit formula: p^ = exp(f)/[1+exp(f)]. These estimates were summed over all pixels to obtain an estimate for the prevalence of *S. humilis* in the study area.

All statistical analyses were performed using R version 2.9.2 (R Development Core Team 2009).

## Results

### Prediction model

The Maxent habitat suitability model for *Scorzonera humilis* shown in Fig. 2 obtained a resubstitution AUC value of 0.964. The prediction map showed RPPO values for the species in areas dominated by agricultural land and close to roads in rural regions. This accorded with the relative contributions of predictor variables to the Maxent model (Table 2): 46.9 % for Distance to nearest road and 38.5 % for distance to arable land. Land-cover classes and altitude contributed 7.1 and 5.7 % to the model, respectively, while the contributions of the remaining four variables were negligible.

### Field validation

*Scorzonera humilis* was observed in 14 (2.7 %) of the 528 sample plots in the evaluation data set. The observed frequency of occurrence (OFO) in the evaluation data increased as a function of the relative predicted probabilities of occurrence (RPPO), as shown both by the decile plot (Fig. 3) and, most notably, by the logistic regression model [logit(TPP) = 4.207·RPPO – 5.798], the slope of which was strongly significantly different from zero ($p = 0.00017$; Fig. 4). The prevalence of *S. humilis* in the study area was estimated to be 0.007, the species was thus predicted to be present in 0.7 % of the 3 462 010 (25 m × 25 m) pixels.

Evaluation of the Maxent model by the independent evaluation data set resulted in an AUC value of 0.810 (Fig. 5). According to the MDT and MST criteria, the optimal RPPO thresholds were 0.714 and 0.786, respectively. The sensitivity and specificity of the best model according to the MST criterion were 0.769 and 0.728, in that order.

The extended evaluation data set of observed presences recorded during collection of the evaluation showed strong concentration of observations to high RPPO classes, although *S. humilis* was observed over the entire RPPO range (Fig. 6). The low number of patches observed in pixels with RPPO > 90 % accords with the low overall frequency of pixels in this class. Pixels were distributed on decile RPPO as follows: < 0.1: 2507662; 0.1–0.2: 313050; 0.2–0.3: 170742; 0.3–0.4: 135974; 0.4–0.5: 119571; 0.5–0.6: 99610; 0.6–0.7: 63716; 0.7–0.8: 36108; 0.8–0.9: 14049 and > 0.9: 1528.

## Discussion

### Reliability of the *Scorzonera humilis* HSM model

AUC values in the ranges 0.9–1.0 and 0.8–0.9 are characterised as corresponding to excellent and good models [Pearce and Ferrier (2000); Araújo and Guisan (2006)]. The AUC value for set-aside test data of 0.96 is in the high end of AUC ranges reported in comparable empirical HSM studies (e.g. Guisan et al. 2006b, Hirzel et al. 2006, Wollan et al. 2008), and the AUC value for independent evaluation data of 0.81 is well above the median AUC obtained for HSM models for species in the extensive NCEAS data, grouped by taxonomy, region, grain size or sample size (Elith et al. 2006, Guisan et al. 2007, Wisz et al. 2008). AUC values obtained for the *S. humilis* model are also higher than obtained for comparable models for other rare species associated with agricultural land, such as *Eryngium spinalba* (Marage et al.

2008) and *Arnica montana* (Parolo et al. 2008) in the Alps. Furthermore, AUC values are known to be lower for HSM models when the study area does not extent beyond the edges of the extent of occurrence of the target species (Lobo 2008), and for species with low prevalence (Maggini et al. 2006). Thus, even a cautious interpretation of the AUC results with due consideration of the criticism of the AUC by Lobo et al. (2008) clearly shows that our habitat suitability model for *S. humilis* should be considered as very well fitting (see discussion of evaluation of HSM models below).

Despite our model for *S. humilis* is rated as good to excellent by general evaluation criteria, we are of the opinion that the model is likely to be burdened by many types of errors and uncertainties. These are partly relating to properties of the species such as low detectability (Guisan et al. 2006a) and errors in species occurrence data (e.g. determination errors; Hirzel and Le Lay 2008), small size and biased (spatial) sampling of training data (Wisz et al. 2008, Phillips et al. 2009), uncertainties in GIS analyses (Bakkestuen 2009) and model specification errors, notably overfitting by including too many predictors in the model (Wollan et al. 2008).

Low detectability and determination errors do not apply to our case because *S. humilis* individuals are easily visible when flowering. Exclusive use of geo-referenced presence observations confirmed by own field work also amends shortcomings of herbarium and check-list data related to inaccurate localisation. Nevertheless, the training data set is of the presence-only type and as such it represents a selective and potentially biased sample. Sampling bias is an important source of errors in HSM models, as demonstrated among others by Loiselle et al. (2008). Use of information about the intensity of collection activities when generating pseudo-absences, may improve Maxent models considerably (Phillips et al. 2009). However, because *S. humilis* is easily detected and has been actively searched for by amateur botanists due to its rarity, we believe that museum collection data make up a more representative training data set for modelling the distribution of this species (i.e, a set with lower sampling bias) as compared to many other species. Our modelling results confirm results of previous studies (e.g. Guisan et al. 2006a) that Maxent produces meaningful HSM models.

We think that the fine grain used in the present study of *Scorzonera humilis* is an important reason for high predictive power of the resulting HSM model. Better performance of models generated from finer-scaled than from coarser-scaled (1 km linear scale or coarser) data is also found in studies by Engler et al. (2004) and Guisan et al. (2007). Most likely, 25 m is an appropriate resolution for modelling *S. humilis* habitat suitability because the decisive

factors for occurrence of *the species* (see below), such as fine-scaled topography patterns and land-use (Halvorsen et al. 2009), vary on similar scales.

**The fundamental niche of *S. humilis*: lessons learnt from the HSM model**

Ideally, habitat suitability models attempt at modelling the fundamental niche of a species, i.e. to assess limits for all environmental factors within which a species can survive, grow and reproduce (Begon et al. 1996). In the real world, however, only a subset of the important niche factors can be represented by variables for which full coverage over millions of pixels (the number of pixels in the present study is ~3.5 million) can be achieved. Lack of relevant predictors is one of the main causes of ecologically misleading HSM models and obtaining good predictor variable data is considered one of the main challenges for prediction modelling (Hirzel and Le Lay 2008). We are aware that the environmental predictors available for our HSM modelling by no means represent all niche factors of potential importance for *S. humilis*. Nevertheless, our model satisfies general criteria for a good HSM model. Distance to roads and arable land are by far the most important predictors in our model. We cannot rule out that this is caused by sample selection bias, i.e. overrepresentation of road verges in the training data set, in fact the large difference in AUC value between the set-aside training data set and the independent evaluation data set suggest that the training data is biased in some way. Nevertheless, high performance of the model also as evaluated by the independent evaluation data suggests that these variables are important for the species. This accords with our observations that *S. humilis* is strongly associated with semi-natural grasslands (which occur close to roads and are mostly included in the arable-land category in the land-cover classification) and that the species is able to establish in road verges. Road verges, which are currently subjected to management practices that resemble traditional use of semi-natural grasslands, namely regular cutting (Milberg and Persson 1994), represent the secondmost important habitat for *S. humilis* in the study area.

The two most important predictors of *S. humilis* occurrence are not *environmental* as such, but represent proxies for decisive factors such as the occurrence of a 'traditional' management regime by which semi-natural grasslands are maintained by cutting. The relatively low contribution of land-cover class relative to the distance variables is likely to be due to semi-natural grassland not making up a separate land-cover class.

Observations made in the field (A. Edvardsen, pers. obs.) indicate that *S. humilis* declines rapidly when semi-natural grasslands enter late successional stages characterised by

regrowth after abandonment of management. This indicates that information on former and current use of semi-natural grassland patches could potentially improve the *S. humilis* HSM model, as demonstrated by Marage et al. (2008), who found that information about the current grazing pressure considerably improved a HSM model for *Eryngium spinalba*. Marage et al. (2008) also point to the importance of incorporating historical land-use variables in HSM models. Historical land use is one of the most important sets of niche factors for many species in agricultural landscapes (Cousins and Eriksson 2002). However, historical maps like old cadastral maps are generally not available in digitised form and require considerable processing efforts before they can be used for HSM.

The low contribution of local environmental predictor variables like solar radiation and TWI indicates that these factors are not important for this species. Independent analyses of variation in species composition and environmental factors in *S. humilis* patches in the study area show that the species occurs over a large range of variation in soil reaction and soil moisture (A. Edvardsen et al., unpublished data.). The moderate contribution of altitude reflects almost complete restriction of farms (and semi-natural grasslands) and, hence, *S. humilis*, to lower altitudes in the study area.

**Evaluation of habitat suitability models**

Our results reveal strongly different AUC values by evaluation on set-aside training data (Maxent) and on independently collected evaluation data. Similar results have been obtained for some data sets by Edwards et al. (2006) and Marage et al. (2008). This lends support to the view that evaluation of models should be based upon independent test data from the same region as the training data, collected as an endpoint procedure (Austin 2007). Furthermore, it may suggest a tendency for the present HSM models for *S. humilis* as well as for HSM models in general, to be overfit to the training data (Peterson et al. 2007).

The present study demonstrates that independent evaluation data can be used for several important purposes other than evaluating the HSM model as such. For modelling tools such as Maxent, one important extra benefit is that the prevalence of the species (frequency in the study area) can be estimated and that the relative predicted probability of occurrence (RPPO) values can be re-interpreted in terms of estimated true probabilities of presence (TPP). This is important because performance measures like AUC, as well as the reliability of the model itself, depends on the prevalence of the target species as well as because calibration of RPPO values in terms of TPP strongly increases the usefulness of HSM models (Ward et

al. 2009). Furthermore, data used to evaluate one generation of models can be added to the training data for second-generation models (e.g. Guisan et al. 2006a).

AUC, by far the most widely used measure of HSM performance in comparative studies, has recently been strongly criticised, e.g. by Lobo et al. (2008). Our results confirm that AUC is strongly data-dependent, and that this measure does not provide information on overall model goodness-of-fit when used on presence-only data (Elith and Graham 2009). Comparison of AUC values between species, study areas and data types should be made with great care, taking factors that effect AUC into account. The criticisms that AUC summarises model performance over regions of the ROC space in which one rarely operate, and that omission and commission errors are weighted equally by AUC, are however easily overcome by also reporting other properties of the ROC curve such as MST, MDT and sensitivity and specificity values for different RPPO thresholds. Our results show that for a rare species like *S. humilis*, with low prevalence, the most appropriate threshold according to the MST criterion (the threshold that minimises the overall misclassification rate) slightly favours the ability to predict observed presences (sensitivity ~ 0.8) over the ability to correctly predict observed absences (specificity ~0.7).

**Conclusion: implications for habitat suitability modelling of rare species**

Our Maxent-derived habitat suitability model for *Scorzonera humilis* demonstrates the potential of HSM methods for predicting occurrences also of species that are, at the outset, particularly difficult to predict. For *S. humilis* in SE Østfold, such properties are: (1) rarity (low prevalence), which makes collection of good training (and independent evaluation) data time consuming; (2) scattered occurrence all over the study area due to lack of limitation by climatic factors; (3) affiliation with semi-natural grassland and dependence on traditional management, factors for which relevant descriptors are difficult to obtain, notably with wall-to-wall coverage; and (4) the sparse set of environmental predictors available.

Our results show that independent evaluation data add strongly to the value of HSM models for rare species. Evaluation on independently collected high-quality presence-absence data is not only indispensable for assessing the quality of the niche model, but also by making estimation of true probabilities of occurrence for all pixels in the study area possible. Availability of such estimates opens for stratified probability-based sampling of rare species (and habitat types) by use of spatial predictions from a HSM model (Guisan et al. 2006a).

This is of high relevance for monitoring and assessment studies because sampling costs may be strongly reduced.

Maxent is evidently able to produce good HSM models even for 'difficult' species when the predictor variable set that *a priori* was considered far from optimal. Our result accords with studies demonstrating high ability of the Maxent method to perform well over a wide range of data properties (e.g. Elith et al. 2006). We also believe that the use of a raster for predictor and response variables with a resolution (25 m linear scale) that accords with the size of patches of the target species and the scale of variation of niche factors of importance for this species, contribute strongly to the good performance of the model (cf. also Guisan et al. 2006a, Parolo et al. 2008). Our study does, however, suggest opportunities for further improvement of the *S. humilis* model and for HSM models in general. In particular, we will mention improved land-cover data, availability of digitised historical map information, improved and updated fine-scale resolution maps for derivation of fine-grained predictors of high quality. Improvements in all of these respects will contribute to the making of even more robust habitat suitability models and predictive distribution maps in the future, for rare plant species and other organisms.

**Acknowledgements**

J.I.I. Båtvik, O.W. Krog, S.V. Gustavsen, B.P. Løfall, E. Michaelsen and I. Spikkeland are thanked for providing information on field data. Thanks are also due to J.N. Stokland and S. Mazzoni for participating in the field work planning and for GIS contributions.

**References**

Anonymous. 2008. ArcGIS version 9.3. - ESRI Inc. Redlands, California.

Araújo, M.B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. - J. Biogeogr. 33: 1677-1688.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. - Ecol. Model. 200: 1-19.

Baillie, J.E.M., Hilton-Taylor, C. and Stuart, S.N. 2004. 2004 IUCN Red List of Threatened Species. A Global Species Assessment. - IUCN, Gland and Cambridge.

Bakkestuen, V. 2009. GIS as a tool for studying temporal and spatial patterns in Norwegian ecosystems across disciplines and scales. - Dr. Philos. thesis, University of Oslo.

Begon, M., Harper, J.L. and Townsend, C.R. 1996. Ecology: Individuals, populations and communities, 3rd edn. Blackwell, Oxford.

Bjørdal, I. 2007. Markslagsklassifikasjon i økonomisk kartverk. - Handbok Skog og Landskap: 1-89.
Cousins, S.A.O. and Eriksson, O. 2002. The influence of management history and habitat on plant species richness in a rural hemiboreal landscape, Sweden. - Landsc. Ecol. 17: 517-529.
Edwards, T.C.J., Cutler, D.R., Zimmermann, N.E., Geiser, L. and Alegria, J. 2005 Model-based stratifications for enhancing the detection of rare ecological events. - Ecology 86: 1081-1090.
Edwards, T.C.J., Cutler, D.R., Zimmermann, N.E., Geiser, L. and Moisen, G.G. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. - Ecol. Model. 199: 132-141.
Elith, J., Graham, C.H., Anderson, R.P., Dudık, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Townsend Peterson, A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. and Zimmermann, M.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. - Ecography 29: 129-151.
Elith, J. and Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. - Ecography 32: 66-77.
Engler, R., Guisan, A. and Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. - J. appl. Ecol. 41: 263-274.
Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. - Ecol. Model. 135: 147-186.
Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. and Zimmermann, N.E. 2006. Using niche-based models to improve the sampling of rare species. - Conserv. Biol. 20: 501–511.
Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. and Hastie, T. 2006. Making better geographical predictions of species' distributions. - J. appl. Ecol. 43: 386-392.
Guisan, A., Graham, C.H., Elith, J., Huettmann, F. and Group, N.S.D.M. 2007. Sensitivity of predictive species distribution models to change in grain size. - Divers. Distrib. 13: 332-340.
Haines-Young, R., Barr, C.J., Firbank, L.G., Furse, M., Howard, D.C., McGowan, G., Petit, S., Smart, S.M. and Watkins, J.W. 2003. Changing landscapes, habitats and vegetation diversity across Great Britain. - J. environm. Mgmt. 67: 267-281.

Halvorsen, R., Andersen, T., Blom, H.H., Elvebakk, A., Elven, R., Erikstad, L., Gaarder, G., Moen, A., Mortensen, P.B., Norderhaug, A., Nygaard, K., Thorsnes, T. and Ødegaard, F. 2009. Ny norsk naturtypeinndeling version 1.0, http://www.naturtyper.artsdatabanken.no.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. and Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. - Ecol. Model. 199: 142-152.

Hirzel, A.H. and Le Lay, G. 2008. Habitat suitability modelling and niche theory. - J. appl. Ecol. 45: 1372-1381.

Jaynes, E.T. 1957. Information theory and statistical mechanics. - Phys. Rev. 106: 620-630.

Jiménez-Valverde, A. and Lobo, J. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. - Acta Oecol. 31: 361-369.

Kålås, J.A., Viken, Å. and Bakken, T. 2006. Norsk rødliste 2006 - 2006 Norwegian red list. - Artsdatabanken, Norge, Trondheim.

Lid, J. and Lid, D.T. 2005. Norsk flora. 7 utgåve ved R. Elven. - Det Norske Samlaget, Oslo.

Lobo, J.M. 2008. More complex distribution models or more representative data? - Biodiversity Informatics 5: 14-19.

Lobo, J.M., Jiménez-Valverde, A. and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. - Global Ecol. Biogeogr. 17: 145-151.

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. and Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? - J. Biogeogr. 35: 105-116.

Luoto, M., Rekolainen, S., Aakkula, J. and Pykälä, J. 2003. Loss of plant species richness and habitat connectivity in grasslands associated with agricultural change in Finland. - Ambio 32: 447-452.

Maggini, R., Lehmann, A., Zimmermann, N.E. and Guisan, A. 2006. Improving generalized regression analysis for the spatial prediction of forest communities. - J. Biogeogr. 33: 1729-1749.

Marage, D., Garraud, L. and Rameau, J.-C. 2008. The influence of management history on spatial prediction of *Eryngium spinalba*, an endangered endemic species. - Appl. Veg. Sci. 11: 139-148.

Milberg, P. and Persson, T.S. 1994. Soil seed bank and species recruitment in road verge grassland vegetation. Ann. Bot. Fenn. 31: 155-162.

Moen, A. 1998. National atlas of Norway: vegetation. - Norwegian Mapping Authority, Hønefoss.

Moore, I.D., Grayson, R.B. and Ladson, A.R. 1991. Digital terrain modeling - a review of hydrological, geomorphological and biological applications. - Hydrol. Proc. 5: 3-30.
Økland, R.H. 1989. Hydromorphology and phytogeography of mires in inner Østfold and adjacent part of Akershus, SE Norway, in relation to regional variation in SE Fennoscandian mires. - Opera Bot. 96: 1-122.
Parolo, G., Rossi, G. and Ferrarini, A. 2008. Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. J. Appl. Ecol. 45: 1410-1418.
Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133: 225–245.
Peterson, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. - Q. Rev. Biol. 78: 419-433.
Peterson, A.T., Papes, M. and Eaton, M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. - Ecography 30: 550-560.
Phillips, S.J., Anderson, R.P. and Schapire, R.E. 2006. Maximum entropy modelling of species geographic distributions. - Ecol. Model. 190: 231–259.
Phillips, S.J. and Dudík, M. 2008. Modeling of species distribution with Maxent: new extensions and a comprehensive evaluation. - Ecography 31: 161-175.
Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwich, J.R. and Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. - Ecol. Appl. 19: 181-197.
R Development Core Team. 2009. R: A language and environment for statistical computing. - R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.
Rodrigues, A.S.L., Pilgrim, J.D, Lamoreux, J.F., Hoffmann, M. and Brooks, T.M. 2006. The value of the IUCN Red List for conservation. - Trends Ecol. Evol. 21: 71-76.
Venables, W.N. and Ripley, B.D. 2002. Modern applied statistics with S. - Springer, New York.
Ward, G., Hastie, T., Barry, S., Elith, J. and Leathwick, J.R. 2009. Presence-only data and the EM algorithm. - Biometrics 65: 554-563.
Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. and Group, N.P.S.D.W. 2008. Effects of sample size on the performance of species distribution models. - Divers. Distrib. 14: 763-773.
Wollan, A., Kauserud, H., Gulden, G. and Halvorsen, R. (2008) Modelling and predicting fungal distribution patterns using herbarium data. - J. Biogeogr. 35: 2298-2310.

Table 1. Explanatory variables (7 continuous and 1 categorical) corresponding to all 25 x 25 m pixels in the study area, with summary statistics.

| Variable | Minimum | Maximum | Mean |
|---|---|---|---|
| Distance to nearest road (m) | 0 | 168 | 27 |
| Distance to nearest arable land (m) | 0 | 4606 | 158 |
| Solar radiation, equinox [derived from the digital elevation model (DEM)] | 4788 | 11511 | 9276 |
| Solar radiation, winter (derived from DEM) | 0 | 819 | 326 |
| Solar radiation, summer (derived from DEM) | 23844 | 27783 | 26734 |
| Land-cover class | | Categorical, 37 classes | |
| Topographic wetness index (TWI; derived from DEM) | –3.14 | 8.15 | –0.03 |
| Altitude (derived from DEM), m.a.s.l. | 1 | 208 | 104 |

Table 2. The relative contribution of environmental predictor variables to the Maxent habitat suitability model for *Scorzonera humilis*.

| Predictor variable | Contribution (%) |
|---|---|
| Distance to nearest road | 46.9 |
| Distance to nearest arable land | 38.5 |
| Land-cover class | 7.1 |
| Altitude | 5.7 |
| Solar radiation, equinox | 0.9 |
| Solar radiation, summer | 0.4 |
| Topographic wetness index | 0.3 |
| Solar radiation, winter | 0.2 |

**Figures**

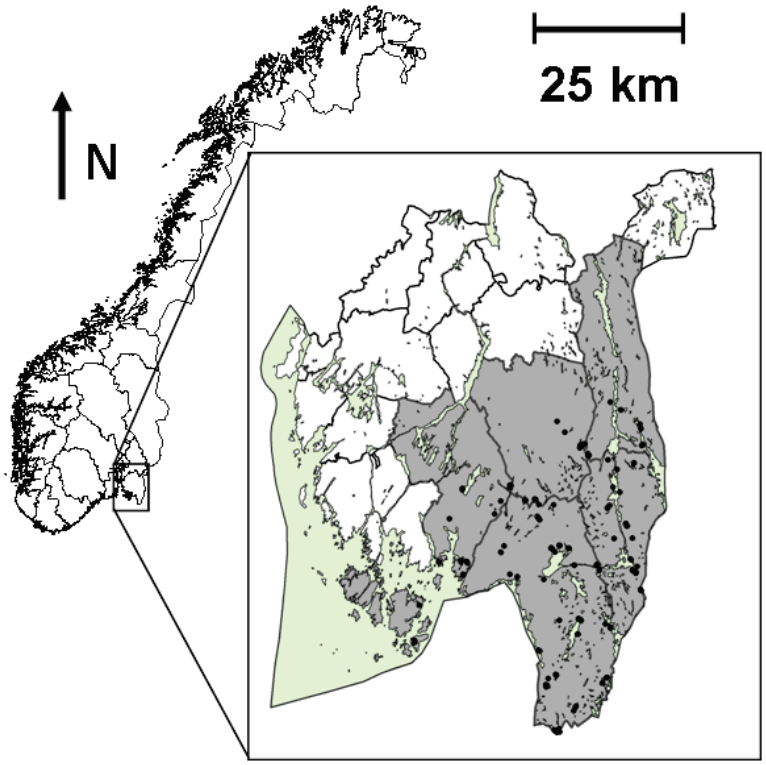


Figure 1. Map showing the 237 study sites in the south-eastern part (grey area) of Østfold county, Norway.

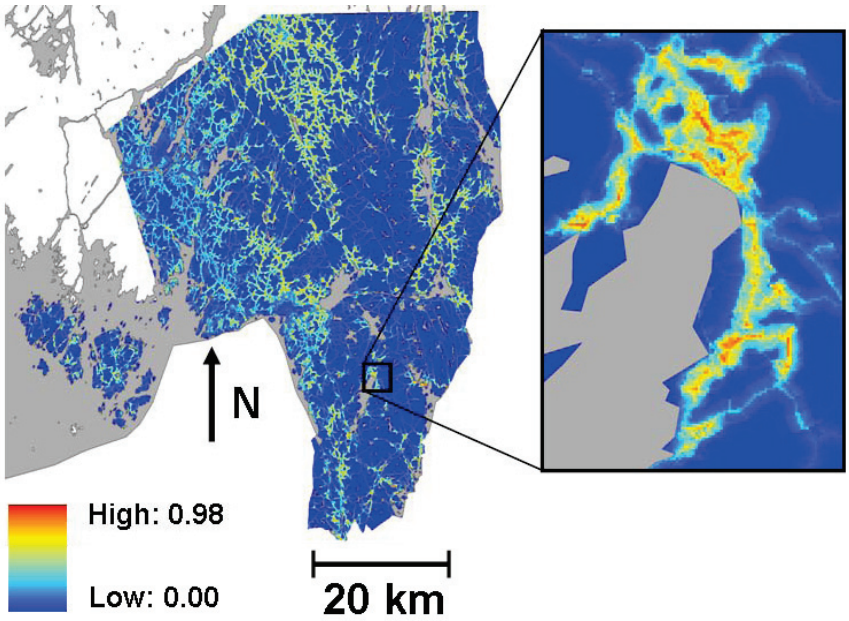


Figure 2. Map showing habitat suitability for *Scorzonera humilis* according to the Maxent model, detailed map inserted.

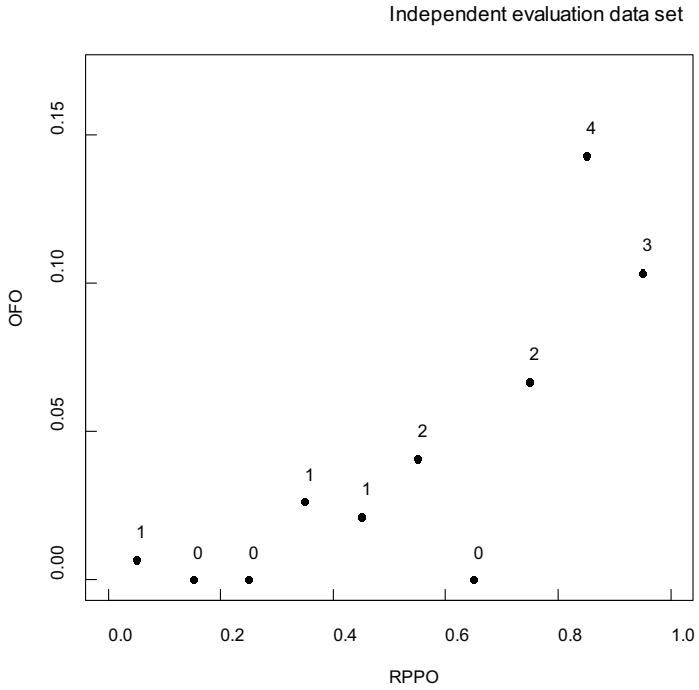


Figure 3. Evaluation of the Maxent HSM model for *Scorzonera humilis* by the independent evaluation data set, plot showing the observed frequency of occurrence (OFO) for each 10-% relative predicted probability of occurrence (RPPO) class. *S. humilis* presence observations detected during collection of the independent evaluation data set are indicated above each point.

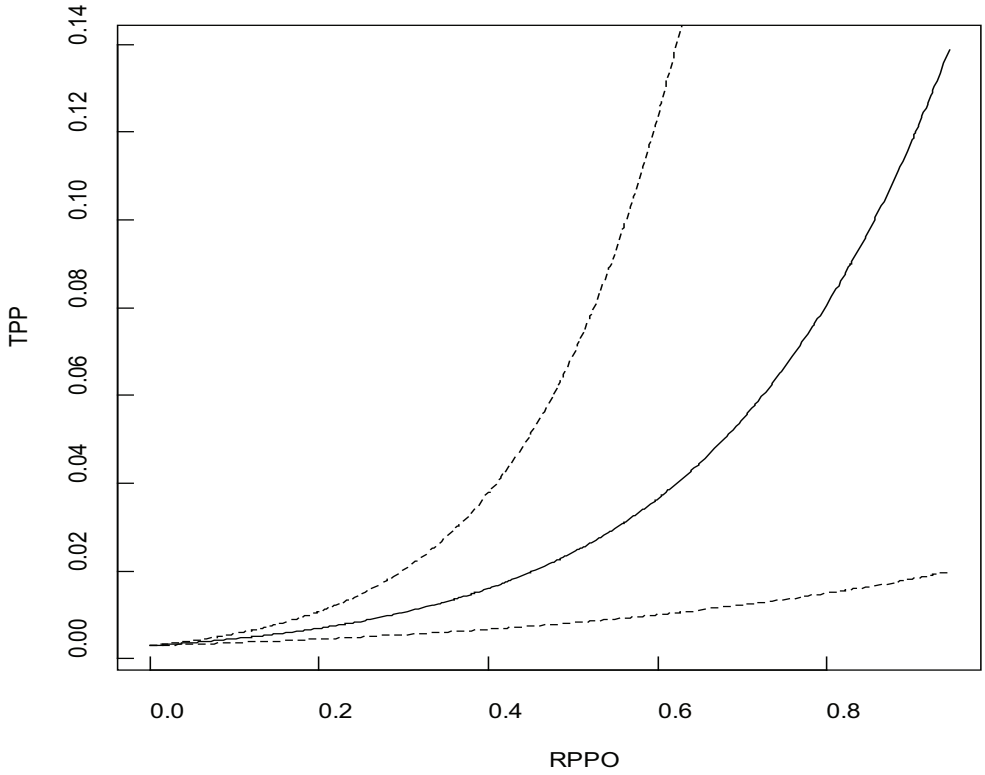


Figure 4. True probability of *Scorzonera humilis* presence (TPP), with 95 % confidence intervals, modelled as a function of relative predicted probability of occurrence (RPPO) by GLM (logistic regression). The slope of the model logit(TPP) = 4.207·RPPO – 5.798 was strongly significantly different from zero (p = 0.00017).

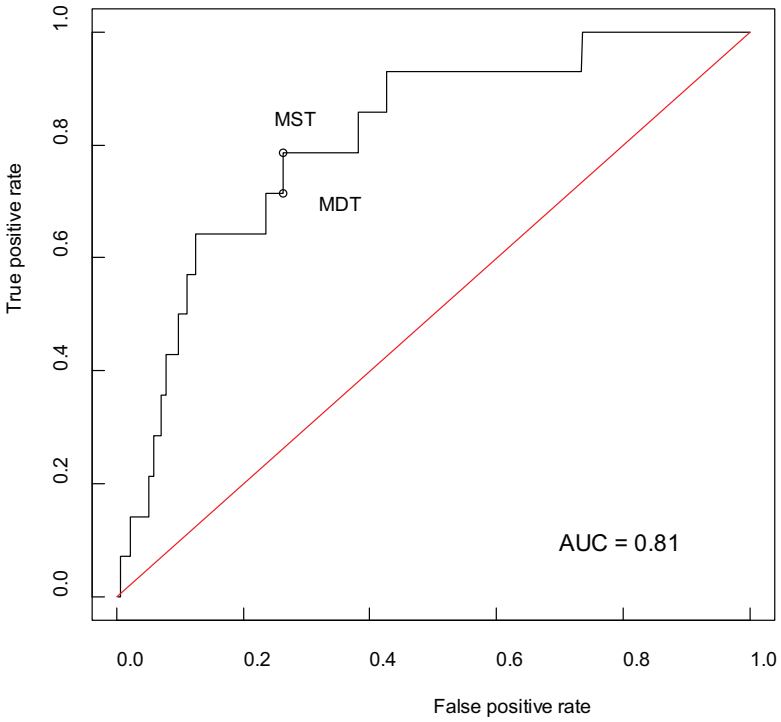


Figure 5. ROC curve for the evaluation data, showing minimum difference threshold (MDT) and maximised sum threshold (MST) values. Sensitivity is synonymous with the True positive rate and specificity equals 1 – False positive rate.

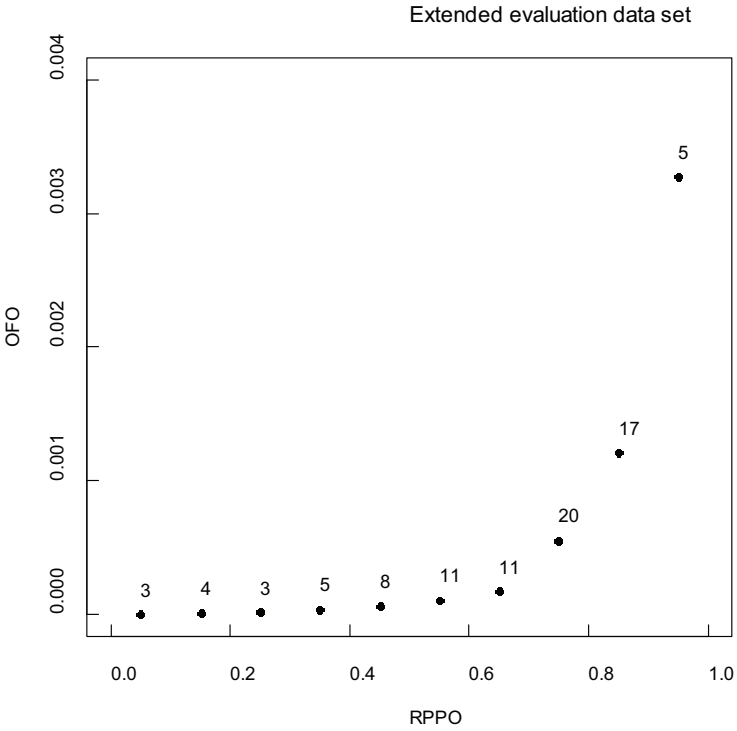


Figure 6. Observed frequency of occurrence (OFO) in the extended evaluation data set distributed on each 10-% relative predicted probability of occurrence (RPPO) class given by Maxent. Number of *S. humilis* patches detected during collection of the extended evaluation data set is indicated above each point.

VI